



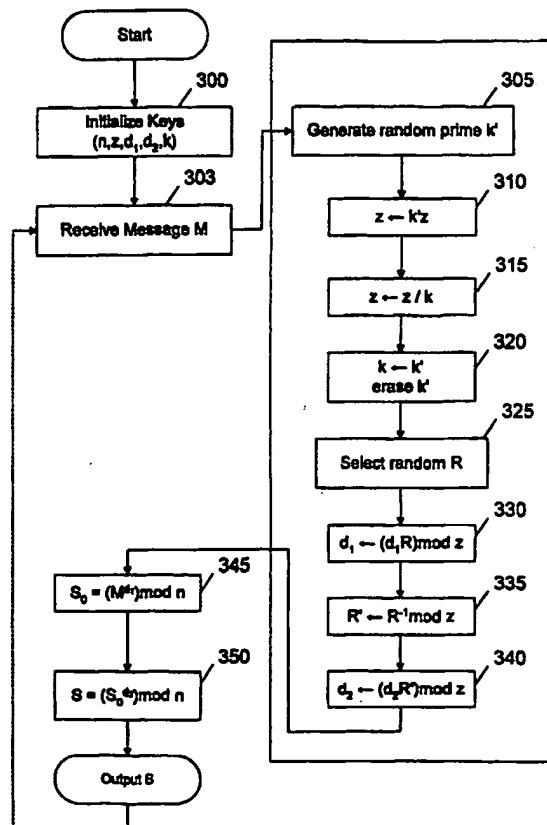
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>H04L 9/30</b>	<b>A1</b>	(11) International Publication Number: <b>WO 99/35782</b> (43) International Publication Date: 15 July 1999 (15.07.99)
<p>(21) International Application Number: PCT/US98/27896</p> <p>(22) International Filing Date: 31 December 1998 (31.12.98)</p> <p>(30) Priority Data: 60/070,344 2 January 1998 (02.01.98) US 60/089,529 15 June 1998 (15.06.98) US</p> <p>(71) Applicant: CRYPTOGRAPHY RESEARCH, INC. [US/US]; Suite 1088, 870 Market Street, San Francisco, CA 94102 (US).</p> <p>(72) Inventors: KOCHER, Paul, C.; 143 Fillmore Street, San Francisco, CA 94117 (US). JAFFE, Joshua, M.; 21B Bird Street, San Francisco, CA 94110 (US).</p> <p>(74) Agents: LAURIE, Ronald, S. et al.; Skadden, Arps, Slate, Meagher &amp; Flom LLP, 525 University Avenue, Palo Alto, CA 94301 (US).</p>	<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>	

(54) Title: LEAK-RESISTANT CRYPTOGRAPHIC METHOD AND APPARATUS

## (57) Abstract

The present invention provides a method and apparatus for securing cryptographic devices against attacks involving external monitoring and analysis. A "self-healing" property is introduced, enabling security to be continually re-established following partial compromises. In addition to producing useful cryptographic results, a typical leak-resistant cryptographic operation modifies or updates (330) secret key material in a manner designed to render useless any information about the secrets that may have previously leaked from the system. Exemplary leak-proof and leak-resistant implementations of the invention are shown for symmetric authentication (350), certified Diffie-Hellman (when either one or both users have certificates), RSA, ElGamal public key decryption (303).



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakhstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

## LEAK-RESISTANT CRYPTOGRAPHIC METHOD AND APPARATUS

This application claims the benefit of US Provisional Application No. 60/070,344 filed January 2, 1998, and US Provisional Application No. 60/089,529 filed June 15, 1998.

### 5 TECHNICAL FIELD

The method and apparatus of the present invention relate generally to cryptographic systems and, more specifically, to securing cryptographic tokens that must maintain the security of secret information in hostile environments.

### BACKGROUND OF THE INVENTION

10 Most cryptosystems require secure key management. In public-key based security systems, private keys must be protected so that attackers cannot use the keys to forge digital signatures, modify data, or decrypt sensitive information. Systems employing symmetric cryptography similarly require that keys be kept secret. Well-designed cryptographic algorithms and protocols should prevent attackers who eavesdrop on communications from  
15 breaking systems. However, cryptographic algorithms and protocols traditionally require that tamper-resistant hardware or other implementation-specific measures prevent attackers from accessing or finding the keys.

If the cryptosystem designer can safely assume that the key management system is completely tamper-proof and will not reveal any information relating to the keys except via  
20 the messages and operations defined in the protocol, then previously known cryptographic techniques are often sufficient for good security. It is currently extremely difficult, however, to make hardware key management systems that provide good security, particularly in low-cost unshielded cryptographic devices for use in applications where attackers will have physical control over the device. For example, cryptographic tokens (such as smartcards used  
25 in electronic cash and copy protection schemes) must protect their keys even in potentially hostile environments. (A token is a device that contains or manipulates cryptographic keys that need to be protected from attackers. Forms in which tokens may be manufactured include, without limitation, smartcards, specialized encryption and key management devices, secure telephones, secure picture phones, secure web servers, consumer electronics devices  
30 using cryptography, secure microprocessors, and other tamper-resistant cryptographic systems.)

A variety of physical techniques for protecting cryptographic devices are known, including enclosing key management systems in physically durable enclosures, coating integrated circuits with special coatings that destroy the chip when removed, and wrapping devices with fine wires that detect tampering. However, these approaches are expensive, difficult to use in single-chip solutions (such as smartcards), and difficult to evaluate since there is no mathematical basis for their security. Physical tamper resistance techniques are also ineffective against some attacks. For example, recent work by Cryptography Research has shown that attackers can non-invasively extract secret keys using careful measurement and analysis of many devices' power consumption. Analysis of timing measurements or electromagnetic radiation can also be used to find secret keys.

Some techniques for hindering external monitoring of cryptographic secrets are known, such as using power supplies with large capacitors to mask fluctuations in power consumption, enclosing devices in well-shielded cases to prevent electromagnetic radiation, message blinding to prevent timing attacks, and buffering of inputs/outputs to prevent signals from leaking out on I/O lines. Shielding, introduction of noise, and other such countermeasures are often, however, of limited value, since skilled attackers can still find keys by amplifying signals and filtering out noise by averaging data collected from many operations. Further, in smartcards and other tamper-resistant chips, these countermeasures are often inapplicable or insufficient due to reliance on external power sources, impracticality of shielding, and other physical constraints. The use of blinding and constant-time mathematical algorithms to prevent timing attacks is also known, but does not prevent more complex attacks such as power consumption analysis (particularly if the system designer cannot perfectly predict what information will be available to an attacker, as is often the case before a device has been physically manufactured and characterized).

The present invention makes use of previously-known cryptographic primitives and operations. For example: U.S. patent 5,136,646 to Haber et al. and the pseudorandom number generator used in the RSAREF cryptographic library use repeated application of hash functions; anonymous digital cash schemes use blinding techniques; zero knowledge protocols use hash functions to mask information; and key splitting and threshold schemes store secrets in multiple parts.

## SUMMARY OF THE INVENTION

The present invention introduces leak-proof and leak-resistant cryptography, mathematical approaches to tamper resistance that support many existing cryptographic primitives, are inexpensive, can be implemented on existing hardware (whether by itself or  
5 via software capable of running on such hardware), and can solve problems involving secrets leaking out of cryptographic devices. Rather than assuming that physical devices will provide perfect security, leak-proof and leak-resistant cryptographic systems may be designed to remain secure even if attackers are able to gather some information about the system and its secrets. This invention describes leak-proof and leak-resistant systems that implement  
10 symmetric authentication, Diffie-Hellman exponential key agreement, ElGamal public key encryption, ElGamal signatures, the Digital Signature Standard, RSA, and other algorithms.

One of the characteristic attributes of a typical leak-proof or leak-resistant cryptosystem is that it is "self-healing" such that the value of information leaked to an attacker decreases or vanishes with time. Leak-proof cryptosystems are able to withstand  
15 leaks of up to  $L_{MAX}$  bits of information per transaction, where  $L_{MAX}$  is a security factor chosen by the system designer to exceed to the maximum anticipated leak rate. The more general class of leak-resistant cryptosystems includes leak-proof cryptosystems, and others that can withstand leaks but are not necessarily defined to withstand any defined maximum information leakage rate. Therefore, any leak-proof system shall also be understood to be  
20 leak-resistant. The leak-resistant systems of the present invention can survive a variety of monitoring and eavesdropping attacks that would break traditional (non-leak-resistant) cryptosystems.

A typical leak-resistant cryptosystem of the present invention consists of three general parts. The initialization or key generation step produces secure keying material appropriate  
25 for the scheme. The update process cryptographically modifies the secret key material in a manner designed to render useless any information about the secrets that may have previously leaked from the system, thus providing security advantages over systems of the background art. The final process performs cryptographic operations, such as producing digital signatures or decrypting messages.

## 30 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows an exemplary leak-resistant symmetric authentication method.

Figure 2 shows an exemplary leak-resistant Diffie-Hellman exponential key exchange operation.

Figure 3 shows an exemplary leak-resistant RSA private key operation.

Figure 4 shows an exemplary leak-resistant ElGamal signing operation.

5

## DETAILED DESCRIPTION OF THE INVENTION

The sections following will describe an introduction to leak-proof/leak-resistant cryptography, followed by various embodiments of the general techniques of the invention as applied to improve the security of common cryptographic protocols.

### 10 I. Introduction and Terminology

The leakage rate  $L$  is defined as the number of bits of useful information about a cryptosystem's secrets that are revealed per operation, where an operation is a cryptographic transaction. Although an attacker may be able to collect more than  $L$  bits worth of measurement data, by definition this data yields no more than  $L$  bits of useful information  
15 about the system's secrets.

The implementer of a leak-proof system chooses a design parameter  $L_{MAX}$ , the maximum amount of leakage per operation the system may allow if it is to remain uncompromised.  $L_{MAX}$  should be chosen conservatively, and normally should significantly exceed the amount of useful information known to be leaked to attackers about the system's  
20 secrets during each transaction. Designers do not necessarily need to know accurately or completely the quantity and type of information that may leak from their systems; the choice of  $L_{MAX}$  may be made using estimates and models for the system's behavior. General factors affecting the choice of  $L_{MAX}$  include the types of monitoring potentially available to attackers, the amount of error in attackers' measurements, and engineering constraints that limit  $L_{MAX}$ .  
25 (Larger values of  $L_{MAX}$  increase memory and performance requirements of the device, and in some cases may increase  $L$ .) To estimate the amount of useful information an attacker could collect by monitoring a device's power consumption, for example, a designer might consider the amount of noise in the device's power usage, the power line capacitance, the useful time resolution for power consumption measurements, as well as the strength of the signals being  
30 monitored. Similarly, the designer knows that timing measurements can rarely yield more than a few bits of information per operation, since timing information is normally quantized to an integral number of clock cycles. In choosing  $L_{MAX}$ , the designer should assume that

attackers will be able to combine information gleaned from multiple types of attacks. If the leakage rate is too large (as in the extreme case where  $L$  equals the key size because the entire key can be extracted during a single transaction), additional design features should be added to reduce  $L$  and reduce the value needed for  $L_{\text{MAX}}$ . Such additional measures can include  
 5 known methods, such as filtering the device's power inputs, adding shielding, introducing noise into the timing or power consumption, implementing constant-time and constant execution path algorithms, and changing the device layout. Again, note that the designer of a leak-resistant system does not actually need to know what information is being revealed or how it is leaked; all he or she need do is choose an upper bound for the rate at which attackers  
 10 might learn information about the keys. In contrast, the designer of a traditional system faces the much harder task of ensuring that no information about the secrets will leak out.

There are many ways information about secrets can leak from cryptosystems. For example, an attacker can use a high-speed analog-to-digital converter to record a smartcard's power consumption during a cryptographic operation. The amount of useful information that  
 15 can be gained from such a measurement varies, but it would be fairly typical to gain enough information to guess each of 128 key bits correctly with a probability of 0.7. This information can reduce the amount of effort required for a brute force attack. For example, a brute force attack with one message against a key containing  $k$  bits where each bit's value is known with probability  $p$  can be completed in

$$20 \quad E(k, p) = \sum_{i=0}^k \left[ \binom{k}{i} (1-p)^i p^{k-i} \left[ \left( \sum_{j=0}^i \binom{k}{j} \right) - \frac{1}{2} \binom{k}{i} \right] + \frac{1}{2} \right]$$

operations. The reduction in the effort for a brute force attack is equivalent to shortening the key by  $L = \log_2(E(k, 1/2) / E(k, p)) = \log_2(k - E(k, p) - 1)$  bits. (For example, in the case of  $k = 128$  and  $p = 0.7$ ,  $L$  is estimated to be about 11 bits for the first measurement. With a multiple message attack, the attacker's effort can fall to as low as  $E(k, p) = \frac{1}{p^k}$ .) Attackers can gain  
 25 additional information about the keys by measuring additional operations; unless leak-resistance is used, finding the key becomes easy after just a few dozen operations.

When choosing  $L_{\text{MAX}}$ , a system designer should consider the signal-to-noise ratio of an attacker's measurements. For example, if the signal and noise are of roughly equivalent magnitude, the designer knows that an attacker's measurements should be incorrect about 25  
 30 percent of the time (e.g.,  $p = 0.75$  if only one observation per key bit is possible). Many

measurement techniques, such as those involving timing, may have signal-to-noise ratios of 1:100 or worse. With such systems,  $L$  is generally quite small, but attackers who can make a large number of measurements can use averaging or other statistical techniques to recover the entire key. In extreme cases, attackers may be able to obtain all key bits with virtually perfect accuracy from a single transaction (i.e.,  $L = k$ ), necessitating the addition of shielding, noise  
5 in the power consumption (or elsewhere), and other measures to reduce  $p$  and  $L$ . Of course,  $L_{\text{MAX}}$  should be chosen conservatively; in the example above where less than 4 useful bits are obtained per operation for the given attack, the designer might select  $L_{\text{MAX}} = 64$  for a leak-proof design.

10 Leak-proof (and, more generally, leak-resistant) cryptosystems provide system designers with important advantages. When designing a traditional (i.e., non-leak-resistant and non-leak-proof) cryptosystem, a careful cryptosystem designer should study all possible information available to attackers if he or she is to ensure that no analytical techniques could be used to compromise the keys. In practice, many insecure systems are developed and  
15 deployed because such analysis is incomplete, too difficult even to attempt, or because the cryptographers working on the system do not understand or cannot completely control the physical characteristics of the device they are designing. Unexpected manufacturing defects or process changes, alterations made to the product by attackers, or modifications made to the product in the field can also introduce problems. Even a system designed and analyzed with  
20 great care can be broken if new or improved data collection and analysis techniques are found later. In contrast, with leak-proof cryptography, the system designer only needs to define an upper bound on the maximum rate at which attackers can extract information about the keys. A detailed understanding of the information available to attackers is not required, since leak-proof (and leak-resistant) cryptosystem designs allow for secret information in the device to  
25 leak out in (virtually) any way, yet remain secure despite this because leaked information is only of momentary value.

In a typical leak-proof design, with each new cryptographic operation  $i$ , the attacker is assumed to be able to choose any function  $F_i$  and determine the  $L_{\text{MAX}}$ -bit result of computing  $F_i$  on the device's secrets, inputs, intermediates, and outputs over the course of the operation.  
30 The attacker is even allowed to choose a new function  $F_i$  with each new operation. The system may be considered leak-proof with a security factor  $n$  and leak rate  $L_{\text{MAX}}$  if, after observing a large number of operations, an attacker cannot forge signatures, decrypt data, or



perform other sensitive operations without performing an exhaustive search to find an  $n$ -bit key or performing a comparable  $O(2^n)$  operation. In addition to choosing  $L_{\text{MAX}}$ , designers also choose  $n$ , and should select a value large enough to make exhaustive search infeasible. In the sections that follow, various embodiments of the invention, as applied to improve the security of common cryptographic operations and protocols, will be described in more detail.

## II. Symmetric Cryptographic Protocols

### A. Symmetric Authentication

An exemplary cryptographic protocol that can be secured using the techniques of the present invention is symmetric authentication.

#### 1. Conventional Symmetric Authentication

Assume a user wishes to authenticate herself to a server using an  $n$ -bit secret key,  $K$ , known to both the server and the user's cryptographic token, but not known to attackers. The cryptographic token should be able to resist tampering to prevent, for example, attackers from being able to extract secrets from a stolen token. If the user's token has perfect tamper resistance (i.e.,  $L=0$ ), authentication protocols of the background art can be used. Typically the server sends a unique, unpredictable challenge value  $R$  to the user's token, which computes the value  $A = H(R \parallel K)$ , where " $\parallel$ " denotes concatenation and  $H$  is a one-way cryptographic hash function such as SHA. The user sends  $A$  to the server, which independently computes  $A$  (using its copy of  $K$ ) and compares its result with the received value. The user authentication succeeds only if the comparison operation indicates a match.

If the function  $H$  is secure and if  $K$  is sufficiently large to prevent brute force attacks, attackers should not be able to obtain any useful information from the  $(R, A)$  values of old authentication sessions. To ensure that attackers cannot impersonate users by replaying old values of  $A$ , the server generates values of  $R$  that are effectively (with sufficiently high probability) unique. In most cases, the server should also make  $R$  unpredictable to ensure that an attacker with temporary possession of a token cannot compute future values of  $A$ . For example,  $R$  might be a 128-bit number produced using a secure random number generator (or pseudorandom number generator) in the server. The properties of cryptographic hash functions such as  $H$  have been the subject of considerable discussion in the literature, and need not be described in detail here. Hash functions typically provide functionality modeled after a random oracle, deterministically producing a particular output from any input. Ideally, such functions should be collision-resistant, non-invertible, should not leak partial

information about the input from the output, and should not leak information about the output unless the entire input is known. Hash functions can have any output size. For example, MD5 produces 128-bit outputs and SHA produces 160-bit outputs. Hash functions may be constructed from other cryptographic primitives or other hash functions.

5 While the cryptographic security of the protocol using technology of the background art may be good, it is not leak-proof; even a one-bit leak function (with  $L=1$ ) can reveal the key. For example, if the leak function  $F$  equals bit  $(R \bmod n)$  of  $K$ , an attacker can break the system quickly since a new key bit is revealed with every transaction where  $(R \bmod n)$  has a new value. Therefore, there is a need for a leak-proof/leak-resistant symmetric authentication  
10 protocol.

## 2. Leak-Resistant Symmetric Authentication

The following is one embodiment of a leak-resistant (and, in fact, also leak-proof) symmetric authentication protocol, described in the context of a maximum leakage rate of  $L_{\text{MAX}}$  bits per transaction from the token and a security factor  $n$ , meaning that attacks of  
15 complexity  $O(2^n)$ , such as brute-force attacks against an  $n$ -bit key, are acceptable, but there should not be significantly easier attacks. The user's token maintains a counter  $t$ , which is initialized to zero, and an  $(n+2L_{\text{MAX}})$ -bit shared secret  $K_t$ , which is initialized with a secret  $K_0$ . Note that against adversaries performing precomputation attacks based on Hellman's time/memory trade-off, larger values of  $n$  may be in order. Note also that some useful  
20 protocol security features, such as user and/or server identifiers in the hash operation inputs, have been omitted for simplicity in the protocol description. It is also assumed that no leaking will occur from the server. For simplicity in the protocol description, some possible security features (such as user and/or server identifiers in the hash operation inputs) have been omitted, and it is assumed that the server is in a physically secure environment.  
25 However, those skilled in the art will appreciate that the invention is not limited to such assumptions, which have been made as a matter of convenience rather than necessity.

As in the traditional protocol, the server begins the authentication process by generating a unique and unpredictable value  $R$  at step 105. For example,  $R$  might be a 128-bit output from a secure random number generator. At step 110, the server sends  $R$  to the user's  
30 token. At step 112, the token receives  $R$ . At step 115, the token increments its counter  $t$  by computing  $t \leftarrow t + 1$ . At step 120, the token updates  $K_t$  by computing  $K_t \leftarrow H_K(t \parallel K_t)$ , where  $H_K$  is a cryptographic hash function that produces an  $(n+2L_{\text{MAX}})$  bit output from the old value

of  $K_t$  and the (newly incremented) value of  $t$ . Note that in the replacement operations (denoted " $\leftarrow$ "), the token deletes the old values of  $t$  and  $K_t$ , replacing them with the new values. By deleting the old  $K_t$ , the token ensures that future leak functions cannot reveal information about the old (deleted) value. At step 122, the token uses the new values of  $t$  and  $K_t$  to compute an authenticator  $A = H_A(K_t \parallel t \parallel R)$ . At step 125, the token sends both  $t$  and the authenticator  $A$  to the server, which receives them at step 130. At step 135, the server verifies that  $t$  is acceptable (e.g., not too large but larger than the value received in the last successful authentication). If  $t$  is invalid, the server proceeds to step 175. Otherwise, at step 140, the server initializes its loop counter  $i$  to zero and its key register  $K_t'$  to  $K_0$ . At step 145, the server compares  $i$  with the received value of  $t$ , proceeding to step 160 if they are equal. Otherwise, at step 150, the server increments  $i$  by computing  $i \leftarrow i + 1$ . At step 155, the server computes  $K_t' \leftarrow H_K(i \parallel K_t')$ , then proceeds back to step 145. At step 160, the server computes  $A' = H_A(K_t' \parallel t \parallel R)$ . Finally, at step 165, the server compares  $A$  and  $A'$ , where the authentication succeeds at step 170 if they match, or fails at 175 if they do not match.

This design assumes that at the beginning of any transaction the attacker may have  $L_{\text{MAX}}$  bits of useful information about the state of the token (e.g.,  $K_t$ ) that were obtained using the leak function  $F$  in a previous operation. During the transaction, the attacker can gain an additional  $L_{\text{MAX}}$  bits of useful information from the token. If, at any time, any  $2L_{\text{MAX}}$  (or fewer) bits of useful information about the secret are known to the attacker, there are still  $(n+2L_{\text{MAX}})-2L_{\text{MAX}} = n$  or more unknown bits. These  $n$  bits of unknown information ensure that attacks will require  $O(2^n)$  effort, corresponding to the desired security factor. However, the attacker should have no more than  $L_{\text{MAX}}$  bits of useful information about  $K_t$  at the end of the transaction. The property that attackers lose useful information during normal operation of the system is a characteristic of the leak-proof or leak-resistant cryptosystem. In general, this information loss is achieved when the cryptosystem performs operations that convert attackers' useful partial information about the secret into useless information. (Information is considered useless if it gives an attacker nothing better than the ability to test candidate values in an  $O(2^n)$  exhaustive search or other "hard" operation. For example, if exhaustive search of  $X$  is hard and  $H$  is a good hash function,  $H(X)$  is useless information to an attacker trying to find  $X$ .)

Thus, the attacker is assumed to begin with  $L_{\text{MAX}}$  bits of useful information about  $K_t$  before the token's  $K_t \leftarrow H_K(t \parallel K_t)$  computation. (Initial information about anything other

than  $K_i$  is of no value to an attacker because  $K_i$  is the only secret value in the token. The function  $H_K$  and the value of  $i$  are not assumed to be secret.) The attacker's information can be any function of  $K_i$  produced from the previous operation's leaks.

### 5    3.    Security Characteristics of Leak-Proof Systems

The following section provides a technical discussion of the security characteristics of the exemplary leak-proof system described above. The following analysis is provided as an example of how the design can be analyzed, and how a system may be designed using general assumptions about attackers' capabilities. The discussion and assumptions do not necessarily  
10    apply to other embodiments of the invention and should not be construed as limiting the scope or applicability of the invention in any way.

During the course of a transaction, the leak function  $F$  might reveal up to  $L_{MAX}$  information about the system and its secrets. The design assumes that any information contained in the system may be leaked by  $F$ , provided that  $F$  does not reveal useful new  
15    information about values of  $K_i$  that were deleted before the operation started, and  $F$  does not reveal useful information about values of  $K_i$  that will be computed in future operations. These constraints are completely reasonable, since real-world leaks would not reveal information about deleted or not-yet-existent data. (The only way information about future  
20     $K_i$  values could be leaked would be the bizarre case where the leak function itself included, or was somehow derived from, the function  $H_K$ .) In practice, these constraints on  $F$  are academic and of little concern, but they are relevant when constructing proofs to demonstrate the security of a leak-proof system.

If the leak occurs at the beginning of the  $H_K$  computation, it could give the attacker up to  $2L_{MAX}$  bits of useful information about the input value of  $K_i$ . Because  $K_i$  contains  
25     $(2L_{MAX}+n)$  bits of secret information and the attacker may have up to  $2L_{MAX}$  bits of useful information about the initial value of  $K_i$ , there remain at least  $(2L_{MAX}+n)-2L_{MAX} = n$  bits of information in  $K_i$  that are secret. The hash function  $H_K$  effectively mixes up these  $n$  bits to produce a secure new  $K_i$  during each transaction such that the attacker's information about the old  $K_i$  is no longer useful.

30    If the leak occurs at the end of the  $H_K$  computation, it could give an attacker up to  $L_{MAX}$  bits of information about the final value of  $H_K$ , yielding  $L_{MAX}$  bits of information about

the input to the subsequent transaction. This is not a problem, since the design assumes that attackers have up to  $L_{\text{MAX}}$  bits of information about  $K_I$  at the beginning of each transaction.

A third possibility is that the attacker's  $L_{\text{MAX}}$  bits of information might describe intermediates computed during the operation  $H_K$ . However, even if the attacker could obtain  
 5  $L_{\text{MAX}}$  new bits of information about the input to  $H_K$  and also  $L_{\text{MAX}}$  bits of information about the output from  $H_K$ , the system would be secure, since the attacker would never have more than  $2L_{\text{MAX}}$  bits of information about the input  $K_I$  or more than  $L_{\text{MAX}}$  bits of information about the output  $K_I$ . Provided that  $L_{\text{MAX}}$  bits of information from within  $H_K$  cannot reveal more than  
 10  $L_{\text{MAX}}$  bits of information about the input, or more than  $L_{\text{MAX}}$  bits of information about the output, the system will be secure. This will be true unless  $H_K$  somehow compresses the input to form a short intermediate which is expanded to form the output. While hash functions whose internal states are smaller than their outputs should not be used, most cryptographic hash functions are fine.

A fourth possibility is that part or all of the leak could occur during the  $A = H_A(K_I \parallel t \parallel$   
 15  $R)$  calculation. The attacker's total "budget" for observations is  $L_{\text{MAX}}$  bits. If  $L_1$  bits of leak occur during the  $H_K$  computation, an additional  $L_2$  bits of information can leak during the  $A = H_A(K_I \parallel t \parallel R)$  operation, where  $L_2 \leq L_{\text{MAX}} - L_1$ . If the second leak provides information about  $K_I$ , this is no different from leaking information about the result of the  $H_K$  computation; the attacker will still conclude the transaction with no more than  $L_{\text{MAX}}$  bits of information about  
 20  $K_I$  because  $L_1 + L_2 \leq L_{\text{MAX}}$ . However, the second leak could reveal information about  $A$ . To keep  $A$  secure against leaks (to prevent, for example, an attacker from using a leak to capture  $A$  and using  $A$  before the legitimate user can), the size of  $A$  should include an extra  $L_{\text{MAX}}$  bits (to provide security even if  $L_2 = L_{\text{MAX}}$ ). Like  $H_K$ ,  $H_A$  should not leak information about deleted or future values of  $K_I$  that are not used in or produced by the given operation. As with the  
 25 similar assumptions on leaks from  $H_K$ , this limitation is primarily academic and of little practical concern, since real-world leak functions do not reveal information about deleted or not-yet-computed data. However, designers might be cautious when using unusual designs for  $H_A$  that are based on or derived from  $H_K$ , particularly if the operation  $H_A(K_I \parallel t \parallel R)$  could reveal useful information about the result of computing  $H_K(t \parallel K_I)$ .

### 30 B. Other Leak-Resistant Symmetric Schemes

The same basic technique of updating a key ( $K$ ) with each transaction, such that leakage about a key during one transaction does not reveal useful information about a key in a

subsequent (or past) transaction, can be easily extended to other applications besides authentication.

### 1. Symmetric Data Verification

For example and without limitation, leak-resistant symmetric data verification is often  
5 useful where a device needs to support symmetrically-signed code, data, content, or  
parameter updates (all of which will, as a matter of convenience, be denoted as "data" herein).  
In existing systems, a hash or MAC of the data is typically computed using a secret key and  
the data is rejected if computed hash or MAC does not match a value received with the data.  
For example, a MAC may be computed as  $\text{HMAC}(K, \text{data})$ , where HMAC is defined in "RFC  
10 2104, HMAC: Keyed-Hashing for Message Authentication" by H. Krawczyk, M. Bellare,  
and R. Canetti, 1997. Traditional (non-leak-resistant) designs are often vulnerable to attacks  
including power consumption analysis of MAC functions and timing analysis of comparison  
operations.

In an exemplary leak-resistant verification protocol, a verifying device (the "verifier")  
15 maintains a counter  $t$  and a key  $K_t$ , which are initialized (for example at the factory) with  $t \leftarrow 0$   
and  $K_t \leftarrow K_0$ . Before the transaction, the verifier provides  $t$  to the device providing the  
signed data (the "signer"), which also knows  $K_0$ . The signer uses  $t$  to compute  $K_{t+1}'$  (the  
prime indicating a quantity derived by the signer, rather than at the verifier) from  $K_0$  (or  $K_t'$   
or any other available value of  $K_i'$ ), using the relation  $K_i' = H_K(i \parallel K_{i-1}')$ , computes signature  
20  $S' = \text{HMAC}(K_{t+1}', \text{data})$ , and sends  $S'$  plus any other needed information (such as data or  $t$ )  
to the verifier. The verifier confirms that the received value of  $t$  (if any) matches its value of  
 $t$ , and rejects the signature if it does not. If  $t$  matches, the verifier increments  $t$  and updates  $K_t$   
in its nonvolatile memory by computing  $t \leftarrow t + 1$  and  $K_t \leftarrow H_K(t \parallel K_t)$ . In an alternative  
embodiment, if the received value of  $t$  is larger than the internal value but the difference is not  
25 unreasonably large, it may be more appropriate to accept the signature and perform multiple  
updates to  $K_t$  (to catch up with the signer) instead of rejecting the signature outright. Finally,  
the verifier computes  $S = \text{HMAC}(K_t, \text{data})$  and verifies that  $S = S'$ , rejecting the signature if  $S$   
does not equal the value of  $S'$  received with the data.

### 2. Symmetric Encryption

30 Besides authentication and verification, leak-resistant symmetric cryptography can  
also be tailored to a wide variety of applications and environments. For example, if data

encryption is desired instead of authentication, the same techniques as were disclosed above may be used to generate a key  $K_t$  used for encryption rather than verification.

### 3. Variations in Computational Implementation

In the foregoing, various applications were disclosed for the basic technique of updating a key  $K_t$  in accordance with a counter and deleting old key values to ensure that future leakage cannot reveal information about the now-deleted key. Those skilled in the art will realize, however, that the exemplary techniques described above may be modified in various ways without departing from the spirit and scope of the invention. For example, if communications between the device and the server are unreliable (for example if the server uses voice recognition or manual input to receive  $t$  and  $A$ ), then small errors in the signature may be ignored. (One skilled in the art will appreciate that many functions may be used to determine whether a signature corresponds – sufficiently closely -- to its expected value.) In another variation of the basic technique, the order of operations and of data values may be adjusted, or additional steps and parameters may be added, without significantly changing the invention. In another variation, to save on communication bandwidth or memory, the high order bits or digits of  $t$  may not need to be communicated or remembered. In another variation, as a performance optimization, devices need not recompute  $K_t$  from  $K_0$  with each new transaction. For example, when a transaction succeeds, the server can discard  $K_0$  and maintain the validated version of  $K_t$ . In another variation, if bi-directional authentication is required, the protocol can include a step whereby the server can authenticate itself to the user (or user's token) after the user's authentication is complete. In another variation, if the server needs to be secured against leaks as well (as in the case where the role of "server" is played by an ordinary user), it can maintain its own counter  $t$ . In each transaction, the parties agree to use the larger of their two  $t$  values, where the device with the smaller  $t$  value performs extra updates to  $K_t$  to synchronize  $t$ . In an alternate embodiment for devices that contain a clock and a reliable power source (e.g., battery), the update operation may be performed periodically, for example by computing  $K_t \leftarrow H_K(t \parallel K_t)$  once per second. The token uses the current  $K_t$  to compute  $A = H_A(K_t \parallel t \parallel R)$  or, if the token does not have any means for receiving  $R$ , it can output  $A = H_A(K_t)$ . The server can use its clock and local copy of the secret to maintain its own version of  $K_t$ , which it can use to determine whether received values of  $A$  are recent and correct. All of the foregoing show that the method and apparatus of the present invention can be implemented using numerous variations and modifications to

the exemplary embodiments described herein, as would be understood by one skilled in the art.

### III. Asymmetric Cryptographic Protocols

The foregoing illustrates various embodiments of the invention that may be used with symmetric cryptographic protocols. As will be seen below, still other techniques of the present invention may be used in connection with asymmetric cryptographic operations and protocols. While symmetric cryptosystems are sufficient for some applications, asymmetric cryptography is required for many applications. There are several ways leak resistance can be incorporated into public key cryptosystems, but it is often preferable to have as little impact as possible on the overall system architecture. Most of the exemplary designs have thus been chosen to incorporate leak resistance into widely used cryptosystems in a way that only alters the key management device, and does not affect the certification process, certificate format, public key format, or processes for using the public key.

#### A. Certified Diffie-Hellman

Diffie-Hellman exponential key exchange is a widely used asymmetric protocol whereby two parties who do not share a secret key can negotiate a shared secret key. Implementations of Diffie-Hellman can leak information about the secret exponents, enabling attackers to determine the secret keys produced by those implementations. Consequently, a leak-resistant implementation of Diffie-Hellman would be useful. To understand such a leak-resistant implementation, it will be useful to first review a conventional Diffie-Hellman implementation.

##### 1. Conventional Certified Diffie-Hellman

Typical protocols in the background art for performing certified Diffie-Hellman exponential key agreement involve two communicating users (or devices) and a certifying authority (CA). The CA uses an asymmetric signature algorithm (such as DSA) to sign certificates that specify a user's public Diffie-Hellman parameters (the prime  $p$  and generator  $g$ ), public key ( $p^x \bmod g$ , where  $x$  is the user's secret exponent), and auxiliary information (such as the user's identity, a description of privileges granted to the certificate holder, a serial number, expiration date, etc.). Certificates may be verified by anyone with the CA's public signature verification key. To obtain a certificate, user  $U$  typically generates a secret exponent ( $x_u$ ), computes his or her own public key  $y_u = g^{x_u} \bmod p$ , presents  $y_u$  along with any required auxiliary identifying or authenticating information (e.g., a passport) to the CA,



who issues the user a certificate  $C_u$ . Depending on the system,  $p$  and  $g$  may be unique for each user, or they may be system-wide constants (as will be assumed in the following description of Diffie-Hellman using the background art).

Using techniques of the background art, Alice and Bob can use their certificates to  
 5 establish a secure communication channel. They first exchange certificates ( $C_{\text{Alice}}$  and  $C_{\text{Bob}}$ ). Each verifies that the other's certificate is acceptable (e.g., properly formatted, properly signed by a trusted CA, not expired, not revoked, etc.). Because this protocol will assume that  $p$  and  $g$  are constants, they also check that the certificate's  $p$  and  $g$  match the expected values. Alice extracts Bob's public key ( $y_{\text{Bob}}$ ) from  $C_{\text{Bob}}$  and uses her secret exponent ( $x_{\text{Alice}}$ ) to  
 10 compute  $z_{\text{Alice}} = (y_{\text{Bob}})^{x_{\text{Alice}}} \bmod p$ . Bob uses his secret exponent and Alice's public key to compute  $z_{\text{Bob}} = (y_{\text{Alice}})^{x_{\text{Bob}}} \bmod p$ . If everything works correctly,  $z_{\text{Alice}} = z_{\text{Bob}}$ , since:

$$\begin{aligned} z_{\text{Alice}} &= (y_{\text{Bob}})^{x_{\text{Alice}}} \bmod p \\ &= (g^{x_{\text{Bob}}})^{x_{\text{Alice}}} \bmod p \\ &= (g^{x_{\text{Alice}}})^{x_{\text{Bob}}} \bmod p \\ &= (y_{\text{Alice}})^{x_{\text{Bob}}} \bmod p \\ &= z_{\text{Bob}}. \end{aligned}$$

Thus, Alice and Bob have a shared key  $z = z_{\text{Alice}} = z_{\text{Bob}}$ . An attacker who pretends to be  
 15 Alice but does not know her secret exponent ( $x_{\text{Alice}}$ ) will not be able to compute  $z_{\text{Alice}} = (y_{\text{Bob}})^{x_{\text{Alice}}} \bmod p$  correctly. Alice and Bob can positively identify themselves by showing that they correctly found  $z$ . For example, each can compute and send the other the hash of  $z$  concatenated with their own certificate. Once Alice and Bob have verified each other, they can use a symmetric key derived from  $z$  to secure their communications. (For an  
 20 example of a protocol in the background art that uses authenticated Diffie-Hellman, see "The SSL Protocol Version 3.0" by A. Freier, P. Karlton, and P. Kocher, March 1996.)

## 2. Leak-Resistant Certified Diffie-Hellman

A satisfactory leak-resistant public key cryptographic scheme should overcome the problem that, while certification requires the public key be constant, information about the  
 25 corresponding private key should not leak out of the token that contains it. In the symmetric protocol described above, the design assumes that the leak function reveals no useful information about old deleted values of  $K_i$  or about future values of  $K_i$  that have not yet been

computed. Existing public key schemes, however, require that implementations repeatedly perform a consistent, usually deterministic, operation using the private key. For example, in the case of Diffie-Hellman, a leak-resistant token that is compatible with existing protocols and implementations should be able to perform the secret key operation  $y^x \bmod p$ , while  
 5 ensuring that the exponent  $x$  remains secret. The radical reshuffling of the secret provided by the hash function  $H_k$  in the symmetric approach cannot be used because the device should be able to perform the same operation consistently.

The operations used by the token to perform the private key operation are modified to add leak resistance using the following variables:

10	<u>Register</u>	<u>Comment</u>
	$x_1$	First part of the secret key (in nonvolatile updateable memory)
	$x_2$	Second part of the secret key (in nonvolatile updateable memory)
	$g$	The generator (not secret).
15	$p$	The public prime, preferably a strong prime (not secret).

The prime  $p$  and generator  $g$  may be global parameters, or may be specific to individual users or groups of users (or tokens). In either case, the certificate recipient should be able to obtain  $p$  and  $g$  securely, usually as built-in constants or by extracting them from the certificate.

To generate a new secret key, the key generation device (often but not always the  
 20 cryptographic token that will contain the key) first obtains or generates  $p$  and  $g$ , where  $p$  is the prime and  $g$  is a generator mod  $p$ . If  $p$  and  $g$  are not system-wide parameters, algorithms known in the background art for selecting large prime numbers and generators may be used. It is recommended that  $p$  be chosen with  $\frac{p-1}{2}$  also prime, or at least that  $\phi(p)$  not be smooth. (When  $\frac{p-1}{2}$  is not prime, information about  $x_1$  and  $x_2$  modulo small factors of  $\phi(p)$  may be  
 25 leaked, which is why it is preferable that  $\phi(p)$  not be smooth. Note that  $\phi$  denotes Euler's totient function.) Once  $p$  and  $g$  have been chosen, the device generates two random exponents  $x_1$  and  $x_2$ . The lowest-order bit of  $x_1$  and of  $x_2$  is not considered secret, and may be set to 1. Using  $p$ ,  $g$ ,  $x_1$ , and  $x_2$ , the device can then compute its public key as  $g^{x_1 x_2} \bmod p$  and submit it, along with any required identifying information or parameters needed (e.g.,  $p$  and  $g$ ), to the  
 30 CA for certification.

Figure 2 illustrates the process followed by the token to perform private key operations. At step 205, the token obtains the input message  $y$ , its own (non-secret) prime  $p$ , and its own secret key halves ( $x_1$  and  $x_2$ ). If  $x_1$ ,  $x_2$ , and  $p$  are stored in encrypted and/or

authenticated form, they would be decrypted or verified at this point. At this step, the token should verify that  $1 < y < p-1$ . At step 210, the token uses a random number generator (or pseudorandom number generator) to select a random integer  $b_0$ , where  $0 < b_0 < p$ . At step 215, the token computes  $b_1 = b_0^{-1} \bmod p$ . The inverse computation mod  $p$  may be performed

5 using the extended Euclidean algorithm or the formula  $b_1 = b_0^{\phi(p)-1} \bmod p$ . At step 220, the token computes  $b_2 = b_1^{x_1} \bmod p$ . At this point,  $b_1$  is no longer needed; its storage space may be used to store  $b_2$ . Efficient algorithms for computing modular exponentiation, widely known in the art, may be used to complete step 220. Alternatively, when a fast modular exponentiator is available, the computation  $b_2$  may be performed using the relationship

10  $b_2 = b_0^{\phi(p)-x_1} \bmod p$ . At step 225, the token computes  $b_3 = b_2^{x_2} \bmod p$ . At this point,  $b_2$  is no longer needed; its storage space may be used to store  $b_3$ . At step 230, the token computes  $z_0 = b_0 y \bmod p$ . At this point,  $y$  and  $b_0$  are no longer needed; their space may be used to store  $r_1$  (computed at step 235) and  $z_0$ . At step 235, the token uses a random number generator to select a random integer  $r_1$ , where  $0 < r_1 < \phi(p)$  and  $\gcd(r_1, \phi(p)) = 1$ . (If  $\frac{p-1}{2}$  is known to be

15 prime, it is sufficient to verify that  $r_1$  is odd.) At step 240, the token updates  $x_1$  by computing  $x_1 \leftarrow x_1 r_1 \bmod \phi(p)$ . The old value of  $x_1$  is deleted and replaced with the updated value. At step 245, the token computes  $r_2 = (r_1^{-1}) \bmod \phi(p)$ . If  $\frac{p-1}{2}$  is prime, then  $r_2$  can be found using a modular exponentiator and the Chinese Remainder Theorem. Note that  $r_1$  is not needed after this step, so its space may be used to store  $r_2$ . At step 250, the token updates  $x_2$  by

20 computing  $x_2 \leftarrow x_2 r_2 \bmod \phi(p)$ . The old value of  $x_2$  should be deleted and replaced with the updated value. At step 255, the token computes  $z_1 = (z_0)^{x_1} \bmod p$ . Note that  $z_0$  is not needed after this step, so its space may be used to store  $z_1$ . At step 260, the token computes  $z_2 = (z_1)^{x_2} \bmod p$ . Note that  $z_1$  is not needed after this step, so its space may be used to store  $z_2$ . At step 265, the token finds the exponential key exchange result by computing

25  $z = z_2 b_3 \bmod p$ . Finally, at step 270, the token erases and frees any remaining temporary variables.

The process shown in Figure 2 correctly computes  $z = y^x \bmod p$ , where  $x = x_1 x_2 \bmod \phi(p)$ , since:

$$\begin{aligned}
z &= z_2 b_3 \bmod p \\
&= (z_1^{x_1} \bmod p) (b_2^{x_2} \bmod p) \bmod p \\
&= ((z_0^{x_1} \bmod p)^{x_2}) ((b_1^{x_1} \bmod p)^{x_2}) \bmod p \\
&= (b_0 y \bmod p)^{x_1 x_2} (b_0^{-1} \bmod p)^{x_1 x_2} \bmod p \\
&= y^{x_1 x_2} \bmod p \\
&= y^r \bmod p.
\end{aligned}$$

The invention is useful for private key owners communicating with other users (or devices) who have certificates, and also when communicating with users who do not.

If Alice has a certificate and wishes to communicate with Bob who does not have a certificate, the protocol proceeds as follows. Alice sends her certificate ( $C_{\text{Alice}}$ ) to Bob, who receives it and verifies that it is acceptable. Bob extracts  $y_{\text{Alice}}$  (along with  $p_{\text{Alice}}$  and  $g_{\text{Alice}}$ , unless they are system-wide parameters) from  $C_{\text{Alice}}$ . Next, Bob generates a random exponent  $x_{\text{BA}}$ , where  $0 < x_{\text{AB}} < \phi(p_{\text{Alice}})$ . Bob then uses his exponent  $x_{\text{AB}}$  and Alice's parameters to calculate  $y_{\text{BA}} = (g_{\text{Alice}}^{x_{\text{BA}}}) \bmod p_{\text{Alice}}$  and the session key  $z = (y_{\text{Alice}}^{x_{\text{BA}}}) \bmod p_{\text{Alice}}$ . Bob sends  $y_{\text{BA}}$  to Alice, who performs the operation illustrated in Figure 2 to update her internal parameters and derive  $z$  from  $y_{\text{BA}}$ . Alice then proves that she computed  $z$  correctly, for example by sending Bob  $H(z \parallel C_{\text{Alice}})$ . (Alice cannot authenticate Bob because he does not have a certificate. Consequently, she does not necessarily need to verify that he computed  $z$  successfully.) Finally, Alice and Bob can use  $z$  (or, more commonly, a key derived from  $z$ ) to secure their communications.

If both Alice and Bob have certificates, the protocol works as follows. First, Alice and Bob exchange certificates ( $C_{\text{Alice}}$  and  $C_{\text{Bob}}$ ), and each verifies that other's certificate is valid. Alice then extracts the parameters  $p_{\text{Bob}}$ ,  $g_{\text{Bob}}$ , and  $y_{\text{Bob}}$  from  $C_{\text{Bob}}$ , and Bob extracts  $p_{\text{Alice}}$ ,  $g_{\text{Alice}}$ , and  $y_{\text{Alice}}$  from  $C_{\text{Alice}}$ . Alice then generates a random exponent  $x_{\text{AB}}$  where  $0 < x_{\text{AB}} < \phi(p_{\text{Bob}})$ , computes  $y_{\text{AB}} = (g_{\text{Bob}}^{x_{\text{AB}}}) \bmod p_{\text{Bob}}$ , and computes  $z_{\text{AB}} = (y_{\text{Bob}}^{x_{\text{AB}}}) \bmod p_{\text{Bob}}$ . Bob generates a random  $x_{\text{BA}}$  where  $0 < x_{\text{BA}} < \phi(p_{\text{Alice}})$ , computes  $y_{\text{BA}} = (g_{\text{Alice}}^{x_{\text{BA}}}) \bmod p_{\text{Alice}}$ , and computes  $z_{\text{BA}} = (y_{\text{Alice}}^{x_{\text{BA}}}) \bmod p_{\text{Alice}}$ . Bob sends  $y_{\text{BA}}$  to Alice, and Alice sends  $y_{\text{AB}}$  to Bob. Alice and Bob each perform the operation shown in Figure 2, where each uses the prime  $p$  from their own certificate and their own secret exponent halves ( $x_1$  and  $x_2$ ). For the message  $y$  in Figure 2, Alice uses  $y_{\text{BA}}$  (received from Bob), and Bob uses  $y_{\text{AB}}$  (received from Alice). Using the process shown in Figure 2, Alice computes  $z$ . Using  $z$  and  $z_{\text{AB}}$  (computed

previously), she can find a session key  $K$ . This may be done, for example, by using a hash function  $H$  to compute  $K = H(z \parallel z_{AB})$ . The value of  $z$  Bob obtains using the process shown in Figure 2 should equal Alice's  $z_{AB}$ , and Bob's  $z_{BA}$  (computed previously) should equal Alice's  $z$ . If there were no errors or attacks, Bob should thus be able to find  $K$ , e.g., by computing  $K = H(z_{BA} \parallel z)$ . Alice and Bob now share  $K$ . Alice can prove her identity by showing that she computed  $K$  correctly, for example by sending Bob  $H(K \parallel C_{Alice})$ . Bob can prove his identity by sending Alice  $H(K \parallel C_{Bob})$ . Alice and Bob can then secure their communications by encrypting and authenticating using  $K$  or a key derived from  $K$ .

Note that this protocol, like the others, is provided as an example only; many variations and enhancements of the present invention are possible and will be evident to one skilled in the art. For example, certificates may come from a directory, more than two parties can participate in the key agreement, key escrow functionality may be added, the prime modulus  $p$  may be replaced with a composite number, etc. Note also that Alice and Bob as they are called in the protocol are not necessarily people; they would normally be computers, cryptographic devices, etc.

For leak resistance to be effective, attackers should not be able to gain new useful information about the secret variables with each additional operation unless a comparable amount of old useful information is made useless. While the symmetric design is based on the assumption that leaked information will not survive the hash operation  $H_K$ , this design uses multiplication operations mod  $\phi(p)$  to update  $x_1$  and  $x_2$ . The most common variety of leaked information, statistical information about exponent bits, is not of use to attackers in this design, as the exponent update process ( $x_1 \leftarrow x_1 r_1 \bmod \phi(p)$  and  $x_2 \leftarrow x_2 r_2 \bmod \phi(p)$ ) destroys the utility of this information. The only relevant characteristic that survives the update process is that  $x_1 x_2 \bmod \phi(p)$  remains constant, so the system designer should be careful to ensure that the leak function does not reveal information allowing the attacker to find new useful information about  $x_1 x_2 \bmod \phi(p)$ .

There is a modest performance penalty, approximately a factor of four, for the leak-resistant design as described. One way to improve performance is to remove the blinding and unblinding operations, which are often unnecessary. (The blinding operations prevent attackers from correlating input values of  $y$  with the numbers processed by the modular exponentiation operation.) Alternatively or additionally, it is possible to update and reuse

values of  $b_0$ ,  $b_3$ ,  $r_1$ , and  $r_2$  by computing  $b_0 \leftarrow (b_0)^v \bmod p$ ,  $b_3 \leftarrow (b_3)^v \bmod p$ ,  $r_1 \leftarrow (r_1)^w \bmod \phi(p)$ , and  $r_2 \leftarrow (r_2)^w \bmod \phi(p)$ , where  $v$  and  $w$  are fairly short random exponents. Note that the relationship  $b_3 \leftarrow b_0^{-x_1 x_2} \bmod p$  remains true when  $b_0$  and  $b_3$  are both raised to the power  $v \pmod{p}$ . The relationship  $r_2 = (r_1^{-1}) \bmod \phi(p)$  also remains true when  $r_1$  and  $r_2$  are  
 5 exponentiated  $\pmod{\phi(p)}$ . Other parameter update operations may also be used, such as exponentiation with fixed exponents (e.g.,  $v = w = 3$ ), or multiplication with random values and their inverses,  $\bmod p$  and  $\phi(p)$ . The time per transaction with this update process is about half that of the unoptimized leak-resistant implementation, but additional storage is required and care should be taken to ensure that  $b_0$ ,  $b_3$ ,  $r_1$ , and  $r_2$  will not be leaked or otherwise  
 10 compromised.

It should also be noted that with this particular type of certified Diffie-Hellman, the negotiated key is the same every time any given pair of users communicate. Consequently, though the blinding operation performed using  $b_0$  and  $b_3$  does serve to protect the exponents, the result  $K$  can be leaked in the final step or by the system after the process is complete. If  
 15 storage is available, parties could keep track of the values of  $y$  they have received (or their hashes) and reject duplicates. Alternatively, to ensure that a different result is obtained from each negotiation, Alice and Bob can generate and exchange additional exponents,  $w_{\text{Alice}}$  and  $w_{\text{Bob}}$ , for example with  $0 < w < 2^{128}$  (where  $2^{128} \ll p$ ). Alice sets  $y = (y_{\text{BA}})^{w_{\text{Alice}} w_{\text{Bob}}} \bmod p$  instead of just  $y = y_{\text{BA}}$ , and Bob sets  $y = (y_{\text{AB}})^{w_{\text{Bob}} w_{\text{Alice}}} \bmod p$  instead of  $y = y_{\text{AB}}$  before  
 20 performing the operation shown in Figure 2.

## B. Leak-Resistant RSA

Another asymmetric cryptographic protocol is RSA, which is widely used for digital signatures and public key encryption. RSA private key operations rely on secret exponents. If information about these secret exponents leaks from an implementation, its security can be  
 25 compromised. Consequently, a leak-resistant implementation of RSA would be useful.

To give RSA private key operations resistance to leaks, it is possible to divide the secret exponent into two halves such that information about either half is destroyed with each operation. These are two kinds of RSA private key operations. The first, private key signing, involves signing a message with one's own private key to produce a digital signature  
 30 verifiable by anyone with one's corresponding public key. RSA signing operations involve computing  $S = M^d \bmod n$ , where  $M$  is the message,  $S$  is the signature (verifiable using  $M = S^e$

mod  $n$ ),  $d$  is the secret exponent and equals  $e^{-1} \bmod \phi(n)$ , and  $n$  is the modulus and equals  $pq$ , where  $n$  and  $e$  are public and  $p$  and  $q$  are secret primes, and  $\phi$  is Euler's phi function. An RSA public key consists of  $e$  and  $n$ , while an RSA private key consists of  $d$  and  $n$  (or other representations of them). For RSA to be secure,  $d$ ,  $\phi(n)$ ,  $p$ , and  $q$  should all be secret.

- 5        The other RSA operation is decryption, which is used to recover messages encrypted using one's public key. RSA decryption is virtually identical to signing, since the decrypted message  $M$  is recovered from the ciphertext  $C$  by computing  $M = C^d \bmod n$ , where the ciphertext  $C$  was produced by computing  $C = M^e \bmod n$ . Although the following discussion uses variable names from the RSA signing operation, the same techniques may be applied
- 10      similarly to decryption.

- An exemplary leak-resistant scheme for RSA implementations may be constructed as illustrated in Figure 3. At step 300, prior to the commencement of any signing or decryption operations, the device is initialized with (or creates) the public and private keys. The device contains the public modulus  $n$  and the secret key components  $d_1$ ,  $d_2$ , and  $z$ , and  $k$ , where  $k$  is a
- 15      prime number of medium-size (e.g.,  $0 < k < 2^{128}$ ) chosen at random,  $z = k\phi(n)$ ,  $d_1$  is a random number such that  $0 < d_1 < z$  and  $\gcd(d_1, z) = 1$ , and  $d_2 = (e^{-1} \bmod \phi(n))(d_1^{-1} \bmod z) \bmod z$ . In this invention,  $d_1$  and  $d_2$  replace the usual RSA secret exponent  $d$ . Techniques for generating the initial RSA primes (e.g.,  $p$  and  $q$ ) and modulus ( $n$ ) are well known in the background art. At step 305, the device computes a random prime  $k'$  of medium size (e.g.,  $0 < k' < 2^{128}$ ).
- 20      (Algorithms for efficiently generating prime numbers are known in the art.)

- At step 303, the device (token) receives a message  $M$  to sign (or to decrypt). At step 310, the device updates  $z$  by computing  $z \leftarrow k'z$ . At step 315, the device updates  $z$  again by computing  $z \leftarrow z / k$ . (There should be no remainder from this operation, since  $k$  divides  $z$ .) At step 320,  $k$  is replaced with  $k'$  by performing  $k \leftarrow k'$ . Because  $k'$  will not be used in
- 25      subsequent operations, its storage space may be used to hold  $R$  (produced at step 325). At step 325, the device selects a random  $R$  where  $0 < R < z$  and  $\gcd(R, z) = 1$ . At step 330, the device updates  $d_1$  by computing  $d_1 \leftarrow d_1 R \bmod z$ . At step 335, the device finds the inverse of  $R$  by computing  $R' \leftarrow R^{-1} \bmod z$  using, for example, the extended Euclidean algorithm. Note that  $R$  is no longer needed after this step, so its storage space may be erased and used to hold
- 30       $R'$ . At step 340, the device updates  $d_2$  by computing  $d_2 \leftarrow d_2 R' \bmod z$ . At step 345, the device computes  $S_0 = M^{d_1} \bmod n$ , where  $M$  is the input message to be signed (or the message

to be decrypted). Note that  $M$  is no longer needed after this step, so its storage space may be used for  $S_0$ . At step 350, the device computes  $S = S_0^{d_1} \bmod n$ , yielding the final signature (or plaintext if decrypting a message). Leak-resistant RSA has similar security characteristics as normal RSA; standard message padding, post-processing, and key sizes may be used. Public key operations are also performed normally (e.g.,  $M = S^e \bmod n$ ).

A simpler RSA leak resistance scheme may be implemented by splitting the exponent  $d$  into two halves  $d_1$  and  $d_2$  such that  $d_1 + d_2 = d$ . This can be achieved during key generation by choosing  $d_1$  to be a random integer where  $0 \leq d_1 \leq d$ , and choosing  $d_2 \leftarrow d - d_1$ . To perform private key operations, the device needs  $d_1$  and  $d_2$ , but it does not need to contain  $d$ . Prior to each private key operation, the cryptographic device identifies which of  $d_1$  and  $d_2$  is larger. If  $d_1 > d_2$ , then the device computes a random integer  $r$  where  $0 \leq r \leq d_1$ , adds  $r$  to  $d_2$  (i.e.,  $d_2 \leftarrow d_2 + r$ ), and subtracts  $r$  from  $d_1$  (i.e.,  $d_1 \leftarrow d_1 - r$ ). Otherwise, if  $d_1 \leq d_2$ , then the device chooses a random integer  $r$  where  $0 \leq r \leq d_2$ , adds  $r$  to  $d_1$  (i.e.,  $d_1 \leftarrow d_1 + r$ ), and subtracts  $r$  from  $d_2$  (i.e.,  $d_2 \leftarrow d_2 - r$ ). Then, to perform the private key operation on a message  $M$ , the device computes  $s_1 = M^{d_1} \bmod n$ ,  $s_2 = M^{d_2} \bmod n$ , and computes the signature  $S = s_1 s_2 \bmod n$ . While this approach of splitting the exponent into two halves whose sum equals the exponent can also be used with Diffie-Hellman and other cryptosystems, dividing the exponent into the product of two numbers mod  $\phi(p)$  is usually preferable since the assumption that information about  $d_1 + d_2$  will not leak is less conservative than the assumption that information about  $x_1 x_2 \bmod \phi(p)$  will not leak. In the case of RSA, updates mod  $\phi(n)$  cannot be done safely, since  $\phi(n)$  must be kept secret.

When the Chinese Remainder Theorem is required for performance, it is possible to use similar techniques to add leak resistance by maintaining multiples of the secret primes ( $p$  and  $q$ ) that are updated every time (e.g., multiplying by the new multiple then dividing by the old multiple). These techniques also protect the exponents ( $d_p$  and  $d_q$ ) as multiples of their normal values. At the end of the operation, the result  $S$  is corrected to compensate for the adjustments to  $d_p$ ,  $d_q$ ,  $p$ , and  $q$ .

An exemplary embodiment maintains state information consisting of the values  $n$ ,  $B_i$ ,  $B_p$ ,  $k$ ,  $p_k$ ,  $q_k$ ,  $d_{pk}$ ,  $d_{qk}$ ,  $p_{inv}$ , and  $f$ . To convert a traditional RSA CRT private key (consisting of  $p$ ,  $q$ ,  $d_p$ , and  $d_q$  with  $p < q$ ) into the new representation, a random value for  $k$  is chosen, where  $0 < k < 2^{64}$ . The value  $B_i$  is chosen at random where  $0 < B_i < n$ , and  $R_1$  and  $R_2$  are chosen at



random where  $0 < R_1 < 2^{64}$  and  $0 < R_2 < 2^{64}$ . (Of course, constants such as  $2^{64}$  are chosen as example values. It is possible, but not necessary, to place constraints on random numbers, such as requiring that they be prime.) The leak-resistant private key state is then initialized by setting  $n \leftarrow pq$ ,  $B_f \leftarrow B_i^{-d} \bmod n$ ,  $p_k \leftarrow (k)(p)$ ,  $q_k \leftarrow (k)(q)$ ,  $d_{pk} \leftarrow d_p + (R_1)(p) - R_1$ ,  $d_{qk} \leftarrow d_q + (R_2)(q) - R_2$ ,  $p_{inv} \leftarrow k(p^{-1} \bmod q)$ , and  $f \leftarrow 0$ .

To update the system state, first a random value  $\alpha$  may be produced where  $0 < \alpha < 2^{64}$ . Then compute  $p_k \leftarrow ((\alpha)(p_k)) / k$ ,  $q_k \leftarrow ((\alpha)(q_k)) / k$ ,  $p_{inv} \leftarrow ((\alpha)(p_{inv})) / k$ ,  $k \leftarrow \alpha$ . The exponents  $d_{pk}$  and  $d_{qk}$  may be updated by computing  $d_{pk} \leftarrow d_{pk} \pm (R_3 p_k - R_3 k)$  and  $d_{qk} \leftarrow d_{qk} \pm (R_4 q_k - R_4 k)$ , where  $R_3$  and  $R_4$  can be random or constant values (even 1). The blinding factors  $B_i$  and  $B_f$  may be updated by computing  $B_i \leftarrow B_i^2 \bmod n$  and  $B_f \leftarrow B_f^2 \bmod n$ , by computing new blinding factors, by exponentiating with a value other than 2, etc. Update processes should be performed as often as practical, for example before or after each modular exponentiation process. Before the update begins, a failure counter  $f$  is incremented, and when the update completes  $f$  is set to zero. If  $f$  ever exceeds a threshold value indicating too many consecutive failures, the device should temporarily or permanently disable itself. Note that if the update process is interrupted, memory values should not be left in intermediate states. This can be done by using complete reliable memory updates. If the total set of variable changes is too large for a single complete update, it is possible to store  $\alpha$  first then do each variable update reliably which keeping track of how many have been completed.

To perform a private key operation (such as decryption or signing), the input message  $C$  is received by the modular exponentiator. Next, the value is blinded by computing  $C' \leftarrow (C)(B_i) \bmod n$ . The blinded input message is then used to compute modified CRT intermediates by computing  $m_{pk} \leftarrow (C')^{d_{pk}} \bmod p_k$  and  $m_{qk} \leftarrow (C')^{d_{qk}} \bmod q_k$ . Next in the exemplary embodiment, the CRT intermediates are multiplied by  $k$ , e.g.  $m_{pk} \leftarrow (k)(m_{pk}) \bmod p_k$  and  $m_{qk} \leftarrow (k)(m_{qk}) \bmod q_k$ . The CRT difference is then computed as  $m_{pqk} = (m_{pk} [+ q_k] - m_{qk}) \bmod q_k$ , where the addition of  $q_k$  and/or reduction mod  $q_k$  are optional. (The addition of  $q_k$  ensures that the result is non-negative.) The blinded result can be computed as

$$M' = \frac{(m_{pk})k + p_k \left[ \left( \frac{(p_{inv})(m_{pqk})}{k} \right) \bmod q_k \right]}{k^2}, \text{ then the final result } M \text{ is computed as } M = (M')B_f$$

mod  $n$ .

As one of ordinary skill in the art will appreciate, variant forms of the invention are possible. For example, the computational processes can be re-ordered or modified without significantly changing the invention. Some portions (such as the initial and blinding steps) can be skipped. In another example, it is also possible to use multiple blinding factors (for  
 5 example, instead of or in addition to the value  $k$ ).

In some cases, other techniques may also be appropriate. For example, exponent vector codings may be rechosen frequently using, for example, a random number generator. Also, Montgomery arithmetic may be performed mod  $j$  where  $j$  is a value that is changed with each operation (as opposed to traditional Montgomery implementations where  $j$  is constant  
 10 with  $j = 2^k$ ). The foregoing shows that the method and apparatus of the present invention can be implemented using numerous variations and modifications to the exemplary embodiments described herein, as would be known by one skilled in the art.

### C. Leak-Resistant ElGamal Public Key Encryption and Digital Signatures

Still other asymmetric cryptographic protocols that may be improved using the  
 15 techniques of the invention. For example, ElGamal and related cryptosystems are widely used for digital signatures and public key encryption. If information about the secret exponents and parameters leaks from an ElGamal implementation, security can be compromised. Consequently, leak-resistant implementations of ElGamal would be useful.

The private key in the ElGamal public key encryption scheme is a randomly selected  
 20 secret  $a$  where  $1 \leq a \leq p-2$ . The non-secret parameters are a prime  $p$ , a generator  $\alpha$ , and  $\alpha^a \bmod p$ . To encrypt a message  $m$ , one selects a random  $k$  (where  $1 \leq k \leq p-2$ ) and computes the ciphertext  $(\gamma, \delta)$  where  $\gamma = \alpha^k \bmod p$  and  $\delta = m(\alpha^a \bmod p)^k \bmod p$ . Decryption is performed by computing  $m = \delta(\gamma^{p-1-a}) \bmod p$ . (See the Handbook of Applied Cryptography by A. Menezes, P. van Oorschot, and S. Vanstone, 1997, pages 294-298, for a description  
 25 of ElGamal public-key encryption).

To make the ElGamal public-key decryption process leak-resistant, the secret exponent  $(p-1-a)$  is stored in two halves  $a_1$  and  $a_2$ , such that  $a_1 a_2 = (\phi(p)-a) \bmod \phi(p)$ . When generating ElGamal parameters for this leak-resistant implementation, it is recommended, but not required, that  $p$  be chosen with  $\frac{p-1}{2}$  prime so that  $\phi(p)/2$  is prime. The  
 30 variables  $a_1$  and  $a_2$  are normally chosen initially as random integers between 0 and  $\phi(p)$ .

Alternatively, it is possible to generate  $a$  first, then choose  $a_1$  and  $a_2$ , as by selecting  $a_1$  relatively prime to  $\phi(p)$  and computing  $a_2 = (a^{-1} \bmod \phi(p))(a_1^{-1} \bmod \phi(p)) \bmod \phi(p)$ .

Figure 4 illustrates an exemplary leak-resistant ElGamal decryption process. At step 405, the decryption device receives an encrypted message pair  $(\gamma, \delta)$ . At step 410, the device selects a random  $r_1$  where  $1 \leq r_1 < \phi(p)$  and  $\gcd(r_1, \phi(p)) = 1$ . At step 415, the device updates  $a_1$  by computing  $a_1 \leftarrow a_1 r_1 \bmod \phi(p)$ , over-writing the old value of  $a_1$  with the new value. At step 420, the device computes the inverse of  $r_1$  by computing  $r_2 = (r_1)^{-1} \bmod \phi(p)$ . Because  $r_1$  is not used after this step, its storage space may be used to hold  $r_2$ . Note that if  $\frac{p-1}{2}$  is prime, then  $r_2$  may also be found by finding  $r_2' = r_1^{(p-1)/2-2} \bmod \frac{p-1}{2}$ , and using the CRT to find  $r_2 \pmod{p-1}$ . At step 425, the device updates  $a_2$  by computing  $a_2 \leftarrow a_2 r_2 \bmod \phi(p)$ . At step 430, the device begins the private key (decryption) process by computing  $m' = \gamma^{a_1} \bmod p$ . At step 435, the device computes  $m = \delta (m')^{a_2} \bmod p$  and returns the message  $m$ . If verification is successful, the result equals the original message because:

$$\begin{aligned} (\delta (m')^{a_2}) \bmod p &= (m (\alpha^a)^k (\gamma^{a_1} \bmod p)^{r_1}) \bmod p \\ &= (m \alpha^{ak} (\gamma^{a_1 a_2 \bmod \phi(p)}) \bmod p) \bmod p \\ &= (m \alpha^{ak} ((\alpha^k \bmod p)^{-a \bmod \phi(p)}) \bmod p) \bmod p \\ &= (m \alpha^{ak} (\alpha^{-ak}) \bmod p) \bmod p \\ &= m \end{aligned}$$

As with the ElGamal public key encryption scheme, the private key for the ElGamal digital signature scheme is a randomly-selected secret  $a$ , where  $1 \leq a \leq p-2$ . The public key is also similar, consisting of a prime  $p$ , a generator  $\alpha$ , and public parameter  $y$  where  $y = \alpha^a \bmod p$ . To sign a message  $m$ , the private key holder chooses or precomputes a random secret integer  $k$  (where  $1 \leq k \leq p-2$  and  $k$  is relatively prime to  $p-1$ ) and its inverse,  $k^{-1} \bmod \phi(p)$ .

Next, the signer computes the signature  $(r, s)$ , where  $r = \alpha^k \bmod p$ ,  $s = ((k^{-1} \bmod \phi(p)) [H(m) - ar]) \bmod \phi(p)$ , and  $H(m)$  is the hash of the message. Signature verification is performed using the public key  $(p, \alpha, y)$  by verifying that  $1 \leq r < p$  and by verifying that  $y^r r^s \bmod p = \alpha^{H(m)} \bmod p$ .

To make the ElGamal digital signing process leak-resistant, the token containing the private key maintains three persistent variables,  $a_k$ ,  $w$ , and  $r$ . Initially,  $a_k = a$  (the private exponent),  $w = 1$ , and  $r = \alpha$ . When a message  $m$  is to be signed (or during the

precomputation before signing), the token generates a random number  $b$  and its inverse  $b^{-1} \bmod \phi(p)$ , where  $b$  is relatively prime to  $\phi(p)$  and  $0 < b < \phi(p)$ . The token then updates  $a_k$ ,  $w$ , and  $r$  by computing  $a_k \leftarrow (a_k)(b^{-1}) \bmod \phi(p)$ ,  $w \leftarrow (w)(b^{-1}) \bmod \phi(p)$ , and  $r \leftarrow (r^b) \bmod p$ .

The signature  $(r, s)$  is formed from the updated value of  $r$  and  $s$ , where

- 5  $s = (w(H(m) - a_k r)) \bmod \phi(p)$ . Note that  $a_k$ ,  $w$ , and  $r$  are not randomized prior to the first operation, but should be randomized before exposure to possible attack, since otherwise the first operation may leak more information than subsequent ones. It is thus recommended that a dummy signature or parameter update with  $a_k \leftarrow (a_k)(b^{-1}) \bmod \phi(p)$ ,  $w \leftarrow (w)(b^{-1}) \bmod \phi(p)$ , and  $r \leftarrow (r^b) \bmod p$  be performed immediately after key generation. Valid signatures  
10 produced using the exemplary tamper-resistant ElGamal process may be checked using the normal ElGamal signature verification procedure.

- It is also possible to split all or some the ElGamal variables into two halves as part of the leak resistance scheme. In such a variant,  $a$  is replaced with  $a_1$  and  $a_2$ ,  $w$  with  $w_1$  and  $w_2$ , and  $r$  with  $r_1$  and  $r_2$ . It is also possible to reorder the operations by performing, for example,  
15 the parameter updates as a precomputation step prior to receipt of the enciphered message. Other variations and modifications to the exemplary embodiments described herein will be evident to one skilled in the art.

#### D. Leak-Resistant DSA

- Another commonly used asymmetric cryptographic protocol is the Digital Signature  
20 Algorithm (DSA, also known as the Digital Signature Standard, or DSS), which is defined in "Digital Signature Standard (DSS)," Federal Information Processing Standards Publication 186, National Institute of Standards and Technology, May 19, 1994 and described in detail in the Handbook of Applied Cryptography, pages 452 to 454. DSA is widely used for digital signatures. If information about the secret key leaks from a DSA implementation, security  
25 can be compromised. Consequently, leak-resistant implementations of DSA would be useful.

- In non-leak-proof systems, the private key consists of a secret parameter  $a$ , and the public key consists of  $(p, q, \alpha, y)$ , where  $p$  is a large (usually 512 to 1024 bit) prime,  $q$  is a 160-bit prime,  $\alpha$  is a generator of the cyclic group of order  $q \bmod p$ , and  $y = \alpha^a \bmod p$ . To sign a message whose hash is  $H(m)$ , the signer first generates (or precomputes) a random  
30 integer  $k$  and its inverse  $k^{-1} \bmod q$ , where  $0 < k < q$ . The signer then computes the signature  $(r, s)$ , where  $r = (\alpha^k \bmod p) \bmod q$ , and  $s = (k^{-1} \bmod q)(H(m) + ar) \bmod q$ .

In an exemplary embodiment of a leak-resistant DSA signing process, the token containing the private key maintains two variables in nonvolatile memory,  $a_k$  and  $k$ , which are initialized with  $a_k = a$  and  $k = 1$ . When a message  $m$  is to be signed (or during the precomputation before signing), the token generates a random integer  $b$  and its inverse  $b^{-1} \bmod q$ , where  $0 < b < q$ . The token then updates  $a_k$  and  $k$  by computing  $a_k \leftarrow (a_k b^{-1} \bmod q)(k) \bmod q$ , followed by  $k \leftarrow b$ . The signature  $(r, s)$  is formed from the updated values of  $a_k$  and  $k$  by computing  $r = \alpha^k \bmod p$  (which may be reduced mod  $q$ ), and  $s = [(b^{-1}H(m) \bmod q) + (a_k r) \bmod q] \bmod q$ . As indicated, when computing  $s$ ,  $b^{-1}H(m) \bmod q$  and  $(a_k r) \bmod q$  are computed first, then combined mod  $q$ . Note that  $a_k$  and  $k$  should be randomized prior to the first operation, since the first update may leak more information than subsequent updates. It is thus recommended that a dummy signature (or parameter update) be performed immediately after key generation. Valid signatures produced using the leak-resistant DSA process may be checked using the normal DSA signature verification procedure.

#### IV. Other Algorithms and Applications

Still other cryptographic processes can be made leak-proof or leak-resistant, or may be incorporated into leak-resistant cryptosystems. For example, cryptosystems such as those based on elliptic curves (including elliptic curve analogs of other cryptosystems), secret sharing schemes, anonymous electronic cash protocols, threshold signatures schemes, etc. be made leak resistant using the techniques of the present invention.

Implementation details of the schemes described may be adjusted without materially changing the invention, for example by re-ordering operations, inserting steps, substituting equivalent or similar operations, etc. Also, while new keys are normally generated when a new system is produced, it is often possible to add leak resistance retroactively while maintaining or converting existing private keys.

Leak-resistant designs avoid performing repeated mathematical operations using non-changing (static) secret values, since they are likely to leak out. However, in environments where it is possible to implement a simple function (such as an exclusive OR) that does not leak information, it is possible use this function to implement more complex cryptographic operations.

While the exemplary implementations assume that the leak functions can reveal any information present in the system, designers may often safely use the (weaker) assumption

that information not used in a given operation will not be leaked by that operation. Schemes using this weaker assumption may contain a large table of precomputed subkey values, from which a unique or random subset are selected and/or updated for each operation. For example, DES implementations may use indexed permutation lookup tables in which a few  
5 table elements are exchanged with each operation.

While leak resistance provides many advantages, the use of leak resistance by itself cannot guarantee good security. For example, leak-resistant cryptosystems are not inherently secure against error attacks, so operations should be verified. (Changes can even be made to the cryptosystem and/or leak resistance operations to detect errors.) Similarly, leak resistance  
10 by itself does not prevent attacks that extract the entire state out of a device (e.g.,  $L=L_{MAX}$ ). For example, traditional tamper resistance techniques may be required to prevent attackers from staining ROM or EEPROM memory cells and reading the contents under a microscope. Implementers should also be aware of interruption attacks, such as those that involve  
15 disconnecting the power or resetting a device during an operation, to ensure that secrets will not be compromised or that a single leaky operation will not be performed repeatedly. (As a countermeasure, devices can increment a counter in nonvolatile memory prior to each operation, and reset or reduce the counter value when the operation completes successfully. If the number of interrupted operations since the last successful update exceeds a threshold value, the device can disable itself.) Other tamper resistance mechanisms and techniques,  
20 such as the use of fixed-time and fixed-execution path code or implementations for critical operations, may need to be used in conjunction with leak resistance, particularly for systems with a relatively low self-healing rate (e.g.,  $L_{MAX}$  is small).

Leak-resistant algorithms, protocols, and devices may be used in virtually any application requiring cryptographic security and secure key management, including without  
25 limitation: smartcards, electronic cash, electronic payments, funds transfer, remote access, timestamping, certification, certificate validation, secure e-mail, secure facsimile, telecommunications security (voice and data), computer networks, radio and satellite communications, infrared communications, access control, door locks, wireless keys, biometric devices, automobile ignition locks, copy protection devices, payment systems,  
30 systems for controlling the use and payment of copyrighted information, and point of sale terminals.

The foregoing shows that the method and apparatus of the present invention can be implemented using numerous variations and modifications to the exemplary embodiments described herein, as would be known by one skilled in the art. Thus, it is intended that the scope of the present invention be limited only with regard to the claims below.

## WHAT IS CLAIMED IS:

- 1 1. A method for implementing RSA with the Chinese Remainder Theorem for use in a  
2 cryptographic system, with resistance to leakage attacks against said cryptographic  
3 system, comprising the steps of:
  - 4 (a) obtaining a representation of an RSA private key corresponding to an RSA  
5 public key, said private key characterized by secret factors  $p$  and  $q$ ;
  - 6 (b) storing said representation of said private key in a memory;
  - 7 (c) obtaining a message for use in an RSA cryptographic operation;
  - 8 (d) computing a first modulus, corresponding to a multiple of  $p$ , where the value  
9 of said multiple of  $p$  and the value of said multiple of  $p$  divided by  $p$  are both  
10 unknown to an attacker of said cryptographic system;
  - 11 (e) reducing said message modulo said first modulus;
  - 12 (f) performing modular exponentiation on the result of step (e);
  - 13 (g) computing a second modulus, corresponding to a multiple of  $q$ , where the  
14 value of said multiple of  $q$  and the value of said multiple of  $q$  divided by  $q$  are  
15 both unknown to an attacker of said cryptographic system;
  - 16 (h) reducing said message modulo said second modulus;
  - 17 (i) performing modular exponentiation on the result of step (h);
  - 18 (j) combining the results of said steps (e) and (h) to produce a result which, if  
19 operated on with an RSA public key operation using said RSA public key,  
20 yields said message; and
  - 21 (k) repeating steps (c) through (j) a plurality of times using different values for  
22 said multiple of  $p$  and for said multiple of  $q$ .
- 1 2. The method of claim 1 where:
  - 2 (i) said step (b) includes storing an exponent  $d_p$  of said RSA private key in said  
3 memory as a plurality of parameters;
  - 4 (ii) an arithmetic function of at least one of said plurality of parameters is  
5 congruent to  $d_p$ , modulo  $(p-1)$ ;
  - 6 (iii) none of said parameters comprising said stored  $d_p$  is equal to  $d_p$ ;
  - 7 (iv) an exponent used in said step (f) is at least one of said parameters;
  - 8 (v) at least one of said parameters in said memory changes with said repetitions of  
9 said steps (c) through (j).



- 1 3. The method of claim 2 where said plurality of parameters includes a first parameter  
2 equal to said  $d_p$  plus a multiple of  $\phi(p)$ , and also includes a second parameter equal  
3 to a multiple of  $\phi(p)$ , where  $\phi$  denotes Euler's totient function.
- 1 4. The method of claim 1 where the value of said multiple of  $p$  divided by  $p$  is equal to  
2 the value of said multiple of  $q$  divided by  $q$ .
- 1 5. The method of claim 1 where said multiple of  $p$  and said multiple of  $q$  used in said  
2 steps (c) through (j) are updated and modified in said memory after said step (b).
- 1 6. The method of claim 1 performed in a smart card.
- 1 7. The method of claim 1 where at least two of said steps are performed in an order other  
2 than (a) through (k)
- 1 8. A method for implementing RSA for use in a cryptographic system, with resistance to  
2 leakage attacks against said cryptographic system, comprising the steps of:
- 3 (a) obtaining an RSA private key corresponding to an RSA public key, said RSA  
4 public key having an RSA modulus  $n$ ;
- 5 (b) storing said private key in a memory in a form whereby a secret parameter of  
6 said key is stored as an arithmetic combination of  $\phi(x)$  and a first at least one  
7 key masking parameter, where
- 8 (i) an operand  $x$  in said  $\phi(x)$  is an exact multiple of at least one factor of  
9 said modulus  $n$  of said RSA public key; and
- 10 (ii) said first key masking parameter is unknown to an attacker of said  
11 cryptosystem;
- 12 (iii) a representation of said first key masking parameter is stored in said  
13 memory;
- 14 (iv)  $\phi$  denotes Euler's totient function;
- 15 (c) receiving a message;
- 16 (d) deriving an RSA input from said message;
- 17 (e) performing modular exponentiation to raise said RSA input to a power  
18 dependent on said secret parameter, modulo an RSA modulus stored in said  
19 memory, to produce an RSA result such that said RSA result raised to the

- 20 power of the public exponent of said RSA public key, modulo the modulus of  
21 said RSA public key, equals said RSA input;
- 22 (f) updating said secret parameter in said memory by:
- 23 (i) modifying said first key masking parameter to produce a new key  
24 masking parameter, where said modification is performed in a manner  
25 such that an attacker with partial useful information about said first key  
26 masking parameter has less useful information about said new key  
27 masking parameter; and
- 28 (ii) using said new key masking parameter to update said secret parameter  
29 in said memory;
- 30 (g) repeating steps (d) through (f) a plurality of times, where the power used for  
31 each of said modular exponentiation steps (e) is different.

1 9. The method of claim 8 where said operand  $x$  in said  $\phi(x)$  corresponds to said RSA  
2 modulus  $n$  of said RSA public key.

1 10. The method of claim 8 where said operand  $x$  in said  $\phi(x)$  corresponds to a prime  
2 factor of said RSA modulus  $n$  of said RSA public key, and where said modular  
3 exponentiation of said step (e) is performed using the Chinese Remainder Theorem.

1 11. A method for implementing exponential key exchange for use in a cryptographic  
2 system, with resistance to leakage attacks against said cryptographic system,  
3 comprising the steps of:

4 (a) obtaining, and storing in a memory, exponential key exchange parameters  $g$   
5 and  $p$ , and a plurality of secret exponent parameters on which an arithmetic  
6 relationship may be computed to produce an exponent  $x$ ;

7 (b) using a key update transformation to produce a plurality of updated secret  
8 exponent parameters while maintaining said arithmetic relationship  
9 thereamong;

10 (c) receiving a public value  $y$  from a party with whom said key exchange is  
11 desired;

12 (d) using said updated secret exponent parameters to perform a cryptographic  
13 computation yielding an exponential key exchange result  $z = y^x \bmod p$ ;

14 (e) using said result  $z$  to secure an electronic communication with said party; and  
15 (f) performing said steps (b), (c), (d), and (e) in a plurality of transactions.

- 1 12. The method of claim 11 where each of said transactions involves a different said  
2 party.
- 1 13. The method of claim 11 where said arithmetic relationship is such that said  
2 exponential key exchange result is a product of certain of said secret exponent  
3 parameters, both before and after said step (b).
- 1 14. The method of claim 11 where said key update transformation includes choosing a  
2 random key update value  $r$ ; and where said step (b) includes multiplying one of said  
3 secret exponent parameters by  $r$  and another of said secret exponent parameters by an  
4 inverse of  $r$ , said multiplication being performed modulo  $\phi(p)$ , where  $\phi$  is Euler's  
5 totient function.
- 1 15. The method of claim 11 where said key update transformation includes choosing a  
2 random key update value  $r$ ; and where said step (b) includes adding  $r$  to one of said  
3 secret exponent parameters and subtracting  $r$  from another of said secret exponent  
4 parameters.
- 1 16. The method of claim 15 where said secret exponent parameters include two values  $x_1$   
2 and  $x_2$  such that  $x_1 + x_2$  is congruent to  $x$ , modulo  $\phi(p)$ , where  $\phi$  is Euler's totient  
3 function, and where said step of performing said cryptographic computation yielding  
4 said exponential key exchange result includes computing  $z_1 = y^{x_1} \bmod p$ ,  $z_2 = y^{x_2}$   
5  $\bmod p$ , and  $z = z_1 z_2 \bmod p$ .
- 1 17. A cryptographic token configured to perform cryptographic operations using a secret  
2 key in a secure manner, comprising:  
3 (a) an interface configured to receive power from a source external to said token;  
4 (b) a memory containing said secret key;  
5 (c) a processor:  
6 (i) configured to receive said power delivered via said interface;  
7 (ii) configured to perform said processing using said secret key from said  
8 memory;  
9 (d) said token having a power consumption characteristic:  
10 (i) that is externally measurable; and

- 11 (ii) that varies over time in a manner measurably correlated with said  
12 cryptographic operations; and
- 13 (e) a source of unpredictable information usable in said cryptographic operations  
14 to make determination of said secret key infeasible from external  
15 measurements of said power consumption characteristic.
- 1 18. The cryptographic token of claim 17, in the form of a secure microprocessor.
- 1 19. The cryptographic token of claim 17, in the form of a smart card.
- 1 20. The cryptographic token of claim 19, wherein said cryptographic operations  
2 performed by said smart card enable a holder thereof to decrypt an encrypted  
3 communication received via a computer network.
- 1 21. The cryptographic token of claim 19, wherein said smart card is configured to store  
2 value in an electronic cash scheme.
- 1 22. The cryptographic token of claim 21, wherein said cryptographic operations include  
2 authenticating that a balance of said stored value has been decreased.
- 1 23. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 asymmetric private key operations.
- 1 24. The cryptographic token of claim 23 wherein said cryptographic operations include  
2 exponential key agreement operations.
- 1 25. The cryptographic token of claim 23, wherein said cryptographic operations include  
2 DSA signing operations.
- 1 26. The cryptographic token of claim 23, wherein said cryptographic operations include  
2 ElGamal private key operations.
- 1 27. The cryptographic token of claim 23, wherein said asymmetric private key operations  
2 include RSA private key operations.

- 1 28. The cryptographic token of claim 27 wherein said private key operations include  
2 Chinese Remainder Theorem operations.
- 1 29. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 symmetric encryption operations.
- 1 30. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 symmetric decryption operations.
- 1 31. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 symmetric authentication operations using said secret key.
- 1 32. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 authenticating a payment.
- 1 33. The cryptographic token of claim 17, wherein said cryptographic operations include  
2 securing a broadcast communications signal.
- 1 34. The cryptographic token of claim 33, wherein said cryptographic operations include  
2 decrypting a satellite broadcast.
- 1 35. A method for securely managing and using a private key in a computing environment  
2 where information about said private key may leak to attackers, comprising the steps  
3 of:  
4 (a) using a first private key, complementary to a public key, to perform first  
5 asymmetric cryptographic operation;  
6 (b) reading at least a portion of said first private key from a memory;  
7 (c) transforming said read portion of said first private key to produce a second  
8 private key:  
9 (i) said second private key usable to perform a subsequent asymmetric  
10 cryptographic operation in a manner that remains complementary to  
11 said public key, and  
12 (ii) said transformation enabling said asymmetric cryptographic operations  
13 to be performed in a manner such that information leaked during said

14 first asymmetric cryptographic operation does not provide  
15 incrementally useful information about said second private key;  
16 (d) obtaining a datum;  
17 (e) using said second private key to perform said subsequent asymmetric  
18 cryptographic operation on said datum.

1 36. The method of claim 35 where said asymmetric cryptographic operation includes a  
2 digital signing operation.

1 37. The method of claim 36 where said signing operation is an RSA operation.

1 38. The method of claim 36 where said signing operation is an DSA operation.

1 39. The method of claim 36 where said signing operation is an ElGamal operation.

1 40. The method of claim 35 where said asymmetric cryptographic operation includes a  
2 decryption operation.

1 41. The method of claim 40 where said decryption operation is an RSA operation.

1 42. The method of claim 40 where said decryption operation is an ElGamal operation.

1 43. The method of claim 35 where at least two of said steps are performed in an order  
2 different than (a), (b), (c), (d), (e).

1 44. The method of claim 35 further comprising the step, after at least said step (c), of  
2 replacing said private key in said memory with said second private key.

1 45. The method of claim 35, performed in a smart card.

1 46. The method of claim 35, further comprising the steps of: prior to at least said step (c),  
2 incrementing a counter stored in a nonvolatile memory and verifying that said counter  
3 has not exceeded a threshold value; and after at least said step (c) has completed  
4 successfully, decreasing a value of said counter.

- 1 47. A method for performing cryptographic transactions while protecting a stored  
2 cryptographic key against compromise due to leakage attacks, comprising the steps  
3 of:  
4 (a) retrieving a stored private cryptographic key stored in a memory, said stored  
5 key having been used in a previous cryptographic transaction;  
6 (b) using a first cryptographic function to derive from said stored key an updated  
7 key, about which useful information about said stored key obtained through  
8 monitoring of leaked information is effectively uncorrelated to said updated  
9 key;  
10 (c) replacing said stored key in said memory with said updated key;  
11 (d) using an asymmetric cryptographic function, cryptographically processing a  
12 datum with said updated key; and  
13 (e) sending said cryptographically processed datum to an external device having a  
14 public key corresponding to said stored key.
- 1 48. The method of claim 47 where said stored key includes a first plurality of parameters,  
2 and where said updated key includes a second plurality of parameters.
- 1 49. The method of claim 48 where no secret value within said first plurality of parameters  
2 is included within said second plurality of parameters.
- 1 50. The method of claim 49 where said first plurality of parameters is different than said  
2 second plurality of parameters, yet a predetermined relationship among said first  
3 plurality of parameters is also maintained among said second plurality of parameters.
- 1 51. The method of claim 50 where said relationship among said plurality of parameters is  
2 an arithmetic function involving at least two of said plurality of parameters.
- 1 52. The method of claim 51 where said arithmetic function is the sum of said parameters.
- 1 53. The method of claim 51 where said relationship includes a bitwise combination of  
2 said parameters.
- 1 54. The method of claim 53 where said bitwise combination is an exclusive OR.

- 1 55. The method of claim 47 where said step (b) includes using pseudorandomness to  
2 derive said updated key.
- 1 56. A method for implementing a private key operation for an asymmetric cryptographic  
2 system with resistance to leakage attacks against said cryptographic system,  
3 comprising the steps of:  
4 (a) encoding a portion of a private key as at least two component parts, such that  
5 an arithmetic function of said parts yields said portion;  
6 (b) modifying said component parts to produce updated component parts, but  
7 where said arithmetic function of said updated parts still yields said private  
8 key portion;  
9 (c) obtaining a message for use in an asymmetric private key cryptographic  
10 operation;  
11 (d) separately applying said component parts to said message to produce an  
12 intermediate result;  
13 (e) deriving a final result from said intermediate result such that said final result is  
14 a valid result of applying said private key to said message; and  
15 (f) repeating steps (b) through (e) a plurality of times.
- 1 57. The method of claim 56 where said private key portion includes an exponent, and  
2 where said intermediate result represents the result of raising said message to the  
3 power of said exponent, modulo a second key portion.
- 1 58. The method of claim 57 where said private key operation is configured for use with  
2 an RSA cryptosystem.
- 1 59. The method of claim 57 where said private key operation is configured for use with  
2 an ElGamal cryptosystem.
- 1 60. The method of claim 56 where said private key operation is configured for use with a  
2 DSA cryptosystem.
- 1 61. The method of claim 60 where said private key is represented by secret parameters  $a_k$   
2 and  $k$  whose product, modulo a predetermined DSA prime  $q$  for said private key,  
3 yields said private key portion.



- 1 62. The method of claim 56 implemented in a smart card.
- 1 63. The method of claim 56 where said private key is configured for use with an elliptic  
2 curve cryptosystem.
- 1 64. A method for performing cryptographic transactions in a cryptographic token while  
2 protecting a stored cryptographic key against compromise due to leakage attacks,  
3 including the steps of:  
4 (a) retrieving said stored key from a memory;  
5 (b) cryptographically processing said key, to derive an updated key, by executing  
6 a cryptographic update function that:  
7 (i) prevents partial information about said stored key from revealing  
8 useful information about said updated key, and  
9 (ii) also prevents partial information about said updated key from  
10 revealing useful information about said stored key;  
11 (c) replacing said stored key in said memory with said updated key;  
12 (d) performing a cryptographic operation using said updated key; and  
13 (e) repeating steps (a) through (d) a plurality of times.
- 1 65. The method of claim 64 where said cryptographic update function of said step (b)  
2 includes a one-way hash operation.
- 1 66. The method of claim 64 where said cryptographic operation of said step (d) is a  
2 symmetric cryptographic operation; and comprising the further step of sending a  
3 result of said cryptographic operation to a party capable of rederiving said updated  
4 key.
- 1 67. The method of claim 64 further comprising the step, prior to said step (a), of receiving  
2 from a second party a symmetric authentication code and a parameter; and said where  
3 said step (b) includes iterating a cryptographic transformation a number of times  
4 determined from said parameter; and where said step (d) includes performing a  
5 symmetric message authentication code verification operation.

- 6 68. he method of claim 66 where said step (d) of performing said cryptographic operation  
7 includes using said updated key to encrypt a datum.
- 1 69. The method of claim 66 where said updated key contains unpredictable information  
2 such that said updated key is not stored in its entirety anywhere outside of said  
3 cryptographic token; and where the result of said step (d) is independent of said  
4 unpredictable information.
- 1 70. The method of claim 64 where said step (c) of replacing said stored key includes:  
2 (i) explicitly erasing a region of said memory containing said stored key; and  
3 (ii) storing said updated key in said region of memory.
- 1 71. The method of claim 64 performed within a smart card.

FIG. 1

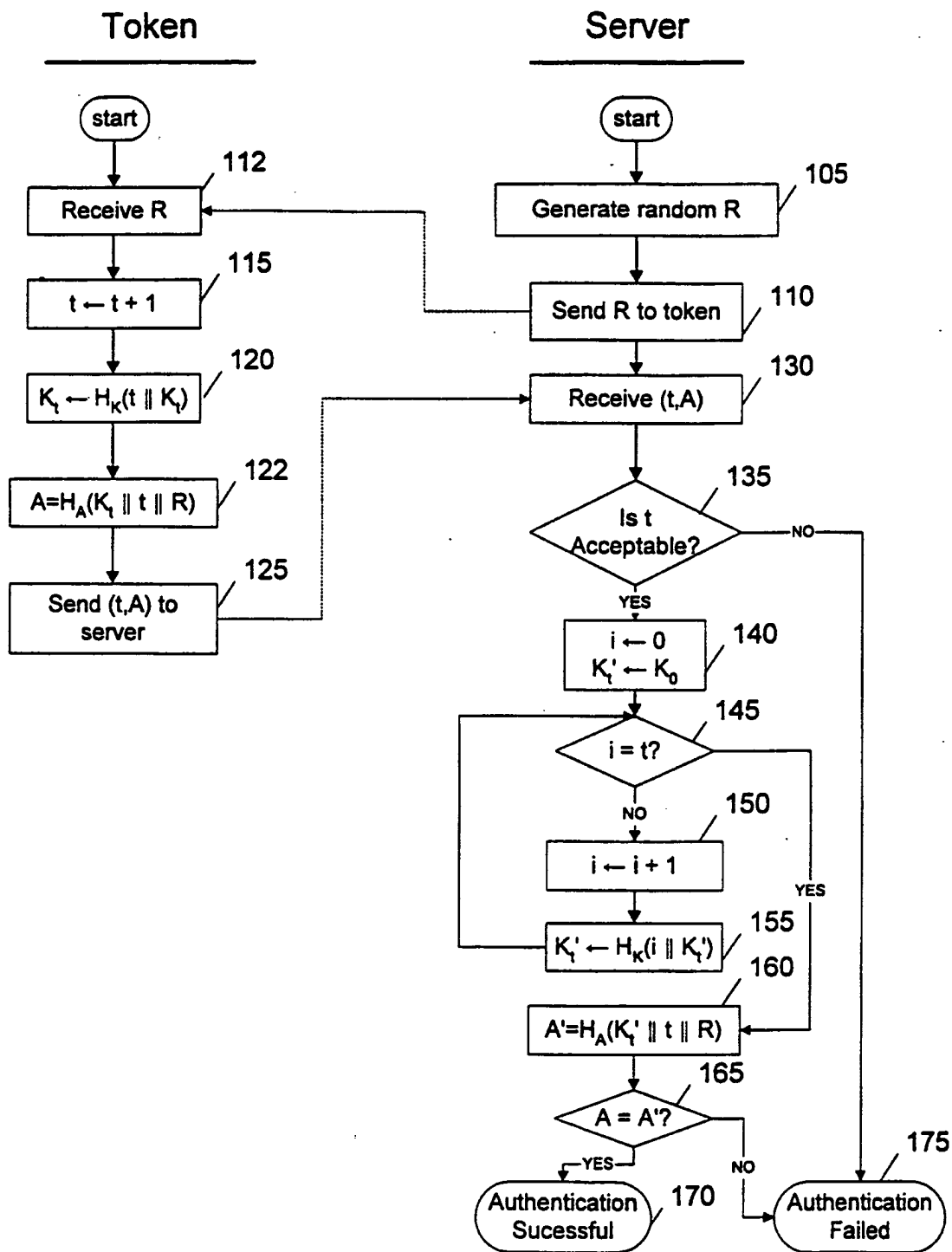


FIG. 2

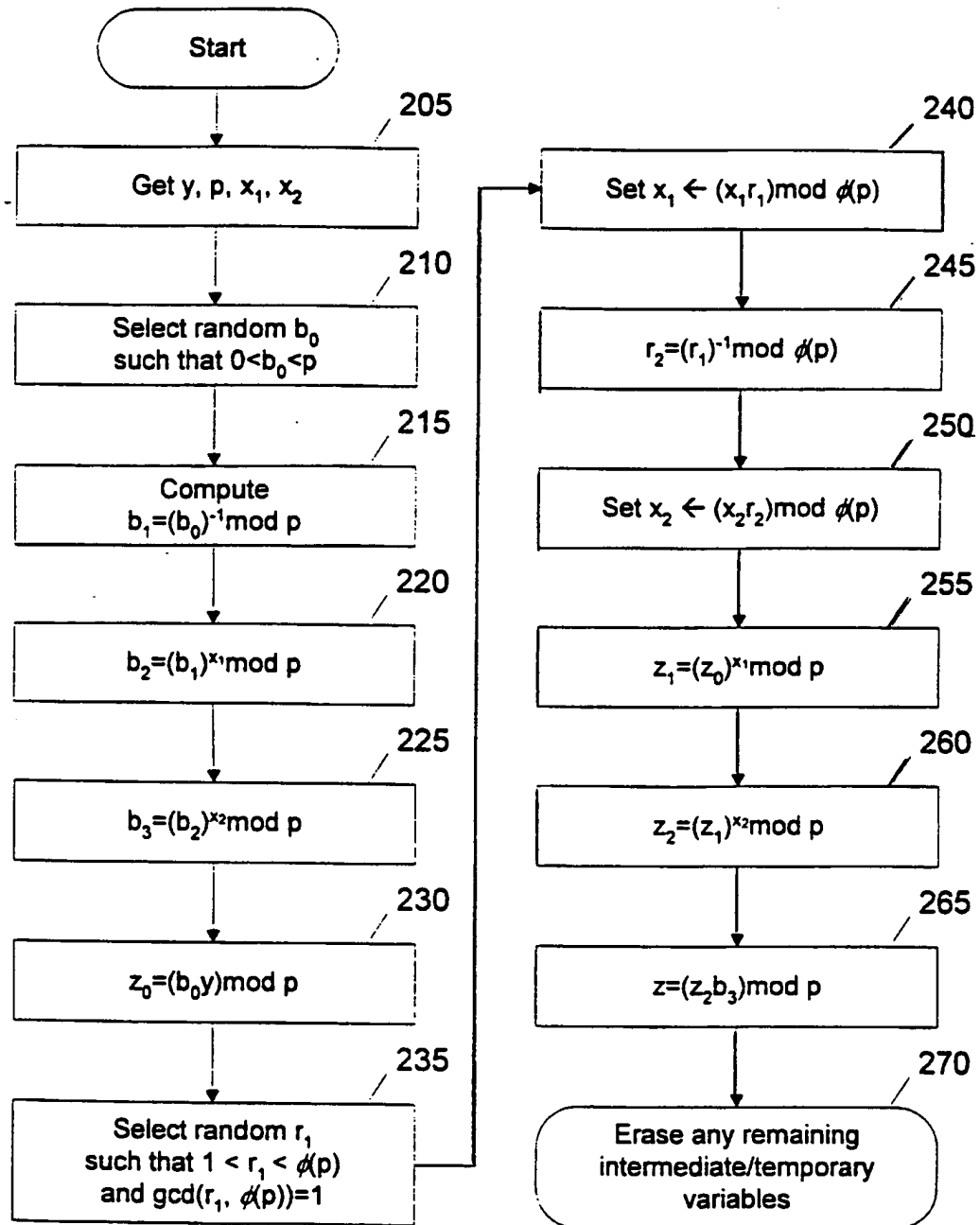


FIG. 3

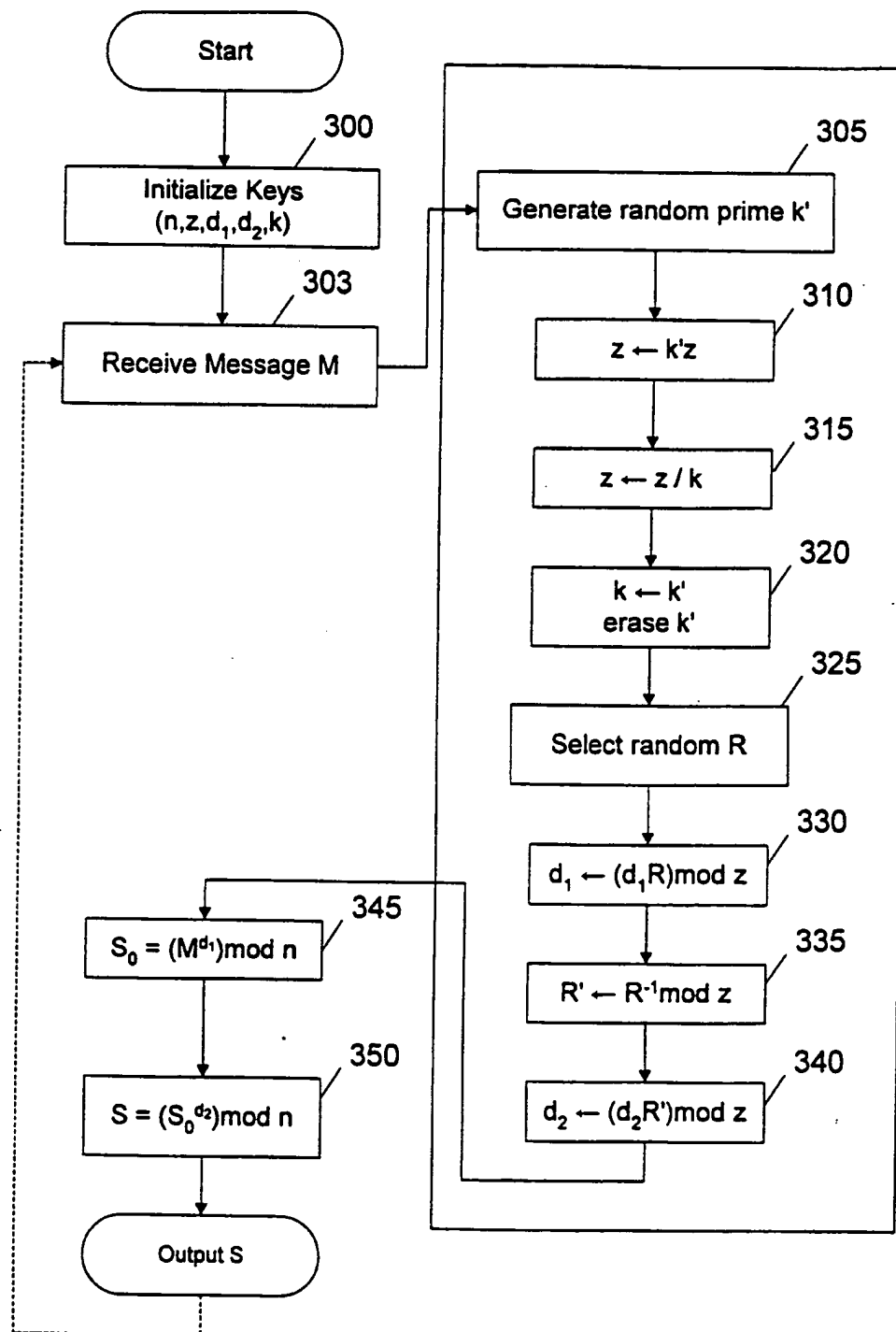
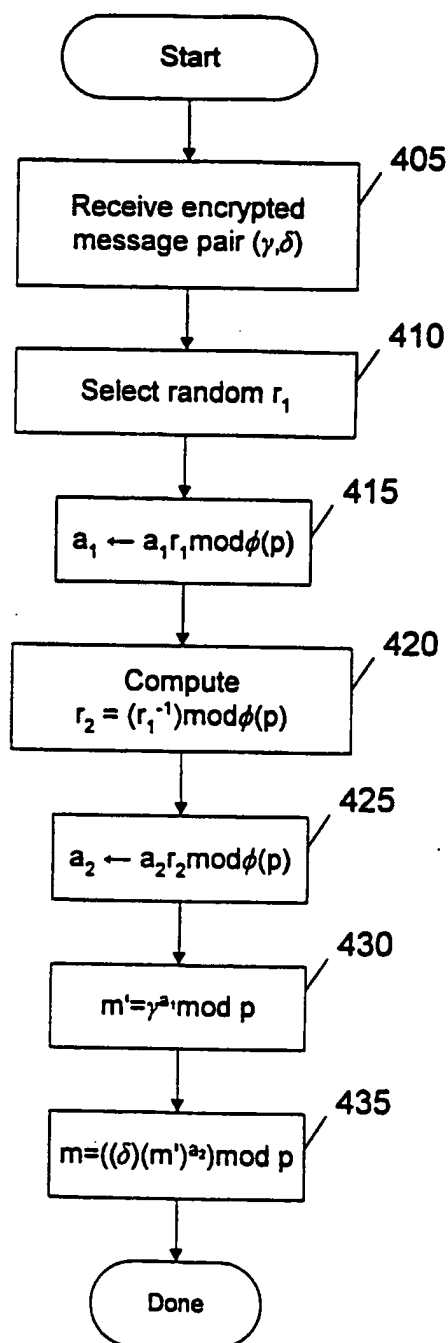


FIG. 4



## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/27896

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : HO4 L 9/30

US CL : 380/30,49

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 380/30,49

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4,799,258 A (DAVIES et al) 17 January 1989, abstract, col.4, lines 43-50, col.7, lines 15-33, col.8, lines 12-19	17-23,25- 45
Y	US 5,546,463 A (CAPUTO et al) 13 August 1996, abstract, col.2, lines, 60-65, col.5, lines 39-50,53-58, col.6, lines 7-12	17-23,25-45
A,P	US 5,848,159 A (COLLINS et al.) 08 December 1998, abstract, col.1, lines 56-67, col.4, lines 33-44, col.5, lines 52-67, col.6, lines 24-30	1-16,46-71



Further documents are listed in the continuation of Box C.



See patent family annex.

*	Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A	document defining the general state of the art which is not considered to be of particular relevance		
*E	earlier document published on or after the international filing date	*X	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*L	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*O	document referring to an oral disclosure, use, exhibition or other means		
*P	document published prior to the international filing date but later than the priority date claimed	*A	document member of the same patent family

Date of the actual completion of the international search

30 MARCH 1999

Date of mailing of the international search report

06 MAY 1999

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-0040

Authorized officer

GAIL HAYES

Telephone No. (703) 305-9711

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/27896

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS

search terms: token, smart card, tamper proof, tamper resistant, leak-resistant, RSA, public key, private key, chinese remainder theorem, diffie hellman, dsa, des